THIS PAGE BLANK (USPTO)

(54) Title: A PROCESS TO STUDY CHANGES IN GENE EXPRESSION IN STEM CELLS

(57) Abstract

    The present invention includes a method to identify stem cell genes that are differentially expressed in stem cells at various stages of
differentiation when compared to undifferentiated stem cells by preparing a gene expression profile of a stem cell population and comparing
the profile to a profile prepared from stem cells at different stages of differentiation, thereby identifying cDNA species, and therefore genes,
which are expressed. The present invention also includes methods to identify a therapeutic agent that modulates the expression of at least
one stem cell gene associated with the differentiation, proliferation and/or survival of stem cells.

-1-

A PROCESS TO STUDY CHANGES
IN GENE EXPRESSION IN STEM CELLS

Technical Field

This invention relates to compositions and methods useful to identify agents that modulate the expression of at least one gene associated with the differentiation, proliferation, dedication and/or survival of stem cells.

5   Background of the Invention

The identification of genes associated with development and differentiation of cells is an important step for advancing our understanding of hematopoiesis, the differentiation of hematopoietic stem cells into erythrocytes, monocytes, platelets and polymorphonuclear white blood cells or granulocytes. The identification of genes

10   associated with hematopoiesis is also an important step for advancing the development of therapeutic agents which modulate, promote or interfere with the differentiation of stem cells.

Hematopoietic stem cells derive from bone marrow stem cells. The bone marrow stem cells ultimately differentiate into the hematopoietic stem cells, which are

15   responsible for the lymphoid, myeloid and erythroid lineages, and stromal stem cells, which differentiate into fibroblasts, osteoblasts, smooth muscle cells, stromal cells and adipocytes (STEWART SELL, IMMUNOLOGY, IMMUNOPATHOLOGY & IMMUNITY, 5th ed. 39-42 Stamford, CT, 1996). The lymphoid lineage, comprising B-cells and T-cells, provides for the production of antibodies, regulation of the cellular immune system, detection of

20   foreign agents in the blood, detection of cells foreign to the host, and the like. The myeloid lineage, which includes monocytes, granulocytes, megakaryocytes as well as others cells, monitors for the presence of foreign bodies in the blood stream, provides protection against neoplastic cells, scavenges foreign materials in the blood stream,

-2-

produces platelets and the like. The erythroid lineage provides the red blood cells which act as oxygen carriers.

Hematopoietic stem cells differentiate as a result from their interaction with growth factors such as interleukins (ILs), lymphokines, colony-stimulating factors

5 (CSFs), erythropoietin (epo), and stem cell factor (SCF). Each of these growth factors have multiple actions that are not necessarily limited to the hematopoietic system (ROBERT A. MEYERS, ED., MOLECULAR BIOLOGY AND BIOTECHNOLOGY: A COMPREHENSIVE DESK REFERENCE, 392-6, New York, 1995). Proliferation, differentiation and survival of immature hematopoietic progenitor cells are sustained by

10 hematopoietic growth factors (hemopoietins). These growth factors also influence the survival and function of mature blood cells. The kinetics of hematopoiesis vary depending on cell type, and their life span may be as little as 6-12 hours to as much as months or years. As a result, the daily renewal of certain lymphocyte progenitors may be substantially lower than that of leukocytic progenitors. The most primitive cells,

15 pluripotent stem cells (PSCs), have high self-renewal capacity (Nathan, 818-821; Saito, *Recent trends in research on differentiation of hematopoietic cells and lymphokines*, Hum. Cell. 5(1): 54 (1992)).

Growth factors are responsible for differentiating the hematopoietic stem cell into either the hemocytoblast, which is the progenitor cell of erythrocytes, neutrophils,

20 eosinophils, basophils, monocytes and platelets, and lymphoid stem cells, which are progenitors to T cells and B cells. SELL, 41. These circulating blood cells are products of terminal differentiation of recognizable precursors (*e.g.*, erythroblasts, mono-myeloblasts and megakaryoblasts, to name but a few). The terminal differentiation of these recognizable precursors may occur exclusively in the marrow cavities of the axial

25 skeleton, with some extension into the proximal femora and humeri (David G. Nathan, *Hematologic Diseases*, IN CECIL TEXTBOOK OF MEDICINE 20th ed., 817, Philadelphia, 1996). White blood cell (WBC) nomenclature may be divided into two major populations on the basis of the form of their nuclei: single nuclei (mononuclear or "round cells") or segmented nuclei (polymorphonuclear).

-3-

In human medicine, the ability to initiate and regulate hematopoiesis is of great importance (McCune *et al.,The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function*, Science 241: 1632(1988)). A variety of diseases and immune disorders, including malignancies, appear to be related to

5   disruptions within the lympho-hematopoietic system. Many of these disorders could be alleviated and/or cured by repopulating the hematopoietic system with progenitor cells, which when triggered to differentiate would overcome the patient's deficiency. In humans, a current replacement therapy is bone marrow transplantation. This type of therapy, however, is both painful (for donor and recipient) because of involvement of

10 invasive procedures and can offer severe complications to the recipient, particularly when the graft is allogeneic and Graft Versus Host Disease (GVHD) results. Therefore, the risk of GVHD restricts the use of bone marrow transplantation to patients with otherwise fatal diseases. A potentially more exciting alternative therapy for hematopoietic disorders is the treatment of patients with reagents that regulate the proliferation and

15 differentiation of stem cells (Lawman *et al.*, U.S. Patent No. 5,650,299 (1997)).

There is also a strong interest in the development of procedures to produce large numbers of the human hematopoietic stem cell. This will allow for identification of growth factors associated with its self regeneration. Additionally, there may be as yet undiscovered growth factors associated (1) with the early steps of dedication of the stem

20 cell to a particular lineage; (2) the prevention of such dedication; and (3) the negative control of stem cell proliferation. Availability of large numbers of stem cells would be extremely useful in bone marrow transplantation, as well as transplantation of other organs in association with the transplantation of bone marrow.

An *in vitro* system that permits determination of what agents induce

25 differentiation or proliferation of progenitor cells within a hematopoietic cell population would have many applications. For example, controlled production of red blood cells would permit the *in vitro* production of red blood cell units for clinical replacement (transfusion) therapy. As is well known, transfused red cells are used in the treatment of anemia following elective surgery, in cases of traumatic blood loss, and in the supportive

30 care of, *e.g.*, cancer patients. Similarly, controlled production of platelets would permit

-4-

the *in vitro* production of platelets for platelet transfusion therapy, which may be used in
cancer patients with thrombocytopenia caused by chemotherapy. For both red cells and
platelets, current volunteer donor pools are accompanied by the risk of infectious
contamination, and availability of an adequate supply can be limited. Determination of
5   such compounds would lend itself to developing methods of controlled *in vitro*
production of specified lineage of mature blood cells to circumvent these problems
(Palsson *et al.*, U.S. Patent No. 5,635,386 (1997)).

       Alternatively, agents could be isolated that selectively deplete a particular lineage
of cells from within a hematopoietic cell population and can similarly confer important
10   advantages. For example, production of stem cells and myeloid cells while selectively
depleting T-cells from a bone marrow cell population could be very important for the
management of patients with human immunodeficiency virus (HIV) infection. Since the
major reservoir of HIV is the pool of mature T-cells, selective eradication of the mature
T-cells from a hematopoietic cell mass collected from a patient has considerable potential
15   therapeutic benefit. If one could selectively remove all the mature T-cells from within an
HIV infected bone marrow cell population while maintaining viable stem cells, the T-cell
depleted bone marrow sample could then be used to "rescue" the patient following
hematolymphoid ablation and autologous bone marrow transplantation. Although there
are reports of the isolation of progenitor cells (see, *e.g.*, Tsukamoto *et al.*, (1991) as
20   representative) such techniques are distinct from the selective removal of T-cells from a
hematopoietic tissue culture (Palsson *et al.*, U.S. Patent No. 5,635,386 (1997)).

Summary of the Invention

       While the differentiation of stem cells has been the subject of intense study, little
is known about the global transcriptional response of stem cells during cell
25   hematopoiesis. The present inventors have devised an approach to systematically assess
the transcriptional regulation of stem cells during hematopoiesis as well as methods for
the identification of agents that modulate the expression of at least one gene associated
with hematopoiesis.

-5-

The present invention includes a method to identify stem cell genes that are differentially expressed in stem cells at various stages of differentiation when compared to undifferentiated stem cells by preparing a gene expression profile of a stem cell population and comparing the profile to a profile prepared from stem cells at different

5   stages of differentiation, thereby identifying cDNA species, and therefore genes, which are expressed.

The present invention further includes a method to identify an agent that modulates the expression of at least one stem cell gene associated with the differentiation process of a stem cell population, comprising the steps of preparing a first gene

10   expression profile of an undifferentiated stem cell population, preparing a second gene expression profile of a stem cell population at a defined stage of differentiation, treating said undifferentiated stem cell population with the agent, preparing a third gene expression profile of the treated stem cell population, and comparing the first, second and third gene expression profiles.   Comparison of the three gene expression profiles for

15   RNA species as represented by cDNA fragments that are differentially expressed upon addition of the agent to the undifferentiated stem cell population identifies agents that modulate the expression of at least one gene in undifferentiated stem cells that is associated with stem cell differentiation.

Another aspect of the invention is a composition comprising a grouping of nucleic

20   acids or nucleic acid fragments affixed to a solid support. The nucleic acids affixed to the solid support correspond to one or more genes whose expression levels are modulated during stem cell differentiation.

Brief Description of the Drawings

Fig. 1   Figure 1 is an autoradiogram of the gene expression profiles generated

25   from cDNAs made with RNA isolated from Lin+, LRH, LRH48 and LRBRH cells. All possible 12 anchoring oligo d(T)n1, n2 were used to generate a complete expression profile for the enzyme ClaI.

-6-

Modes of Carrying Out the Invention

General Description

The differentiation of stem cells during the process of hematopoiesis is a subject of primary importance in view of the need to find ways to modulate the stem cell

5   differentiation process. One means of characterizing the process of hematopoiesis is to measure the ability of stem cells to synthesize specific RNA during stem cell differentiation.

The following discussion presents a general description of the invention as well definitions for certain terms used herein.

10   *Definitions*

The term "stem cells" as used herein, refers to both hematopoietic stem cells and bone marrow stem cells, and includes totipotent cells which serve as progenitors of neoplastic transformation. The term "hematopoietic stem cells" refers to stem cells which differentiate into erythrocytes, monocytes, granulocytes, and platelets. The

15   putative human hematopoietic stem cell may express the cell surface antigen CD34.

The term "hematopoiesis " as used herein, refers to the process by which stem cells differentiate into blood cells, including erythrocytes, monocytes, granulocytes, and platelets.

The term "blood cell", as used herein, refers to all blood cell types derived from the

20   process of hematopoiesis (see STEWART SELL, *IMMUNOLOGY, IMMUNOPATHOLOGY &*
IMMUNITY, 5th ed. 39-42, Stamford, CT, 1996)

The term "solid support", as used herein, refers to any support to which nucleic acids can be bound or immobilized, including nitrocellulose, nylon, glass, other solid supports which are positively charged and nanochannel glass arrays disclosed by Beattie (WO

25   95/1175).

The term "gene expression profile", also referred to as a "differential expression profile" or "expression profile" refers to any representation of the expression level of at

-7-

least one mRNA species in a cell sample or population. For instance, a gene expression profile can refer to an autoradiograph of labeled cDNA fragments produced from total cellular mRNA separated on the basis of size by known procedures. Such procedures include slab gel electrophoresis, capillary gene electrophoresis, high performance liquid chromatography, and the like. Digitized representations of scanned electrophoresis gels are also included as are two and three dimensional representations of the digitized data.

While a gene expression profile encompasses a representation of the expression level of at least one mRNA species, in practice, the typical gene expression profile represents the expression level of multiple mRNA species. For instance, a gene expression profile useful in the methods and compositions disclosed herein represents the expression levels of at least about 5, 10, 20, 50, 100 , 150, 200, 300, 500, 1000 or more preferably, substantially all of the detectable mRNA species in a cell sample or population. Particularly preferred are gene expression profiles or arrays affixed to a solid support that contain a sufficient representative number of mRNA species whose expression levels are modulated under the relevant infection, disease, screening, treatment or other experimental conditions. In some instances a sufficient representative number of such mRNA species will be about 1, 2, 5, 10, 15, 20, 25, 30, 40, 50, 50-75 or 100.

Gene expression profiles can be produced by any means known in the art, including, but not limited to the methods disclosed by: Prashar et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:659-663; Liang et al. (1992) *Science* 257:967-971; Ivanova et al. (1995) *Nucleic Acids Res.* 23:2954-2958; Guilfoyl et al. (1997) *Nucleic Acids Res.* 25(9):1854-1858; Chee et al. (1996) *Science* 274:610-614; Velculescu et al. (1995) *Science* 270:484-487; Fischer et al. (1995) *Proc. Natl. Acad. Sci. USA* 92(12):5331-5335; and Kato (1995) *Nucleic Acids Res.* 23(18):3685-3690.

As an example, gene expression profiles are made to identify one or more genes whose expression levels are modulated during the process of stem cell differentiation. The assaying of the modulation of gene expression via the production of a gene expression profile generally involves the production of cDNA from polyA$^+$ RNA (mRNA) isolated from stem cells as described below.

-8-

Stem cells are harvested or isolated by any technique known in the art. One of the most versatile ways to separate hematopoietic cells is by use of flow cytometry, where the particles, *i.e,.* cells, can be detected by fluorescence or light scattering. The source of the cells may be any source which is convenient. Thus, various tissues, organs, fluids, or

5   the like may be the source of the cellular mixtures. Of particular interest are bone marrow and peripheral blood, although other lymphoid tissues are also of interest, such as spleen, thymus, and lymph node (see Sasaki *et al.,* U.S. Patent No. 5,466,572 and Fei *et al.*, U.S. Patent No. 5,635,387).

Cells of interest will usually be detected and separated by virtue of surface membrane

10   proteins which are characteristic of the cells. For example, CD34 is a marker for immature hematopoietic cells. Markers for dedicated cells may include CD 10, CD19, CD20, and sIg for B cells, CD 15 for granulocytes, CD 16 and CD33 for myeloid cells, CD 14 for monocytes, CD41 for megakaryocytes, CD38 for lineage dedicated cells, CD3, CD4, CD7, CD8 and T cell receptor (TCR) for T cells, Thy-1 for progenitor cells,

15   glycophorin for erythroid progenitors and CD71 for activated T cells. In isolating early progenitors, one may divide a CD34 positive enriched fraction into lineage (Lin) negative, e.g. CD2 - , CD 14 - , CD15 - , CD16 - , CD10 - , CD19 - , CD33 - and glycophorin A - , fractions by negatively selecting for markers expressed on lineage committed cells, Thy-1 positive fractions, or into CD38 negative fractions to provide a

20   composition substantially enriched for early progenitor cells. Other markers of interest include V alpha and V beta chains of the T-cell receptor (Sasaki *et al.,* U. S. Patent No. 5,466,572 (1995)).

After isolation of the appropriate stem cells, total cellular mRNA is isolated from the cell sample. mRNAs are isolated from cells by any one of a variety of techniques.

25   Numerous techniques are well known (*see e..,* Sambrook et al., *Molecular Cloning: A Laboratory Approach*, Cold Spring harbor Press, NY, 1987; Ausbel et., *Current Protocols in Molecular Biology*, Greene Publishing Co. NY, 1995). In general, these techniques first lyse the cells and then enrich for or purify RNA. In one such protocol, cells are lysed in a Tris-buffered solution containing SDS. The lysate is extracted with

30   phenol/chloroform, and nucleic acids precipitated. The mRNAs may be purified from

-9-

- crude preparations of nucleic acids or from total RNA by chromatography, such as

binding and elution from oligo(dT)-cellulose or poly(U)-Sepharose®. However,

purification of poly(A)-containing RNA is not a requirement. As stated above, other

protocols and methods for isolation of RNAs may be substituted.

5   The mRNAs are reverse transcribed using an RNA-directed DNA polymerase, such as

reverse transcriptase isolated from AMV, MoMuLV or recombinantly produced. Many

commercial sources of enzyme are available (*e.g.* Pharmacia, New England Biolabs,

Stratagene Cloning Systems). Suitable buffers., cofactors, and conditions are well known

and supplied by manufacturers (*see also*, Sambrook *et al.* (1989) *Molecular Cloning: a*

10   *laboratory manual*, 2nd Ed., Cold Spring Harbor Laboratory; and Ausbel *et al.*, (1987)

Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience,

N.Y.).

       Various oligonucleotides are used in the production of cDNA. In particular, the

methods utilize oligonucleotide primers for cDNA synthesis, adapters, and primers for

15   amplification. Oligonucleotides are generally synthesized so single strands by standard

chemistry techniques, including automated synthesis. Oligonucleotides are subsequently

de-protected and may be purified by precipitation with ethanol, chromatographed using a

sized or reversed-phase column, denaturing polyacrylamide gel electrophoresis, high-

pressure liquid chromatography (HPLC), or other suitable method. In addition, within

20   certain preferred embodiments, a functional group, such as biotin, is incorporated

preferably at the 5' or 3' terminal nucleotide. A biotinylated oligonucleotide may be

synthesized using pre-coupled nucleotides, or alternatively, biotin may be conjugated to

the oligonucleotide using standard chemical reactions. Other functional groups, such as

florescent dyes, radioactive molecules, digoxigenin, and the like, may also be

25   incorporated.

       Partially-double stranded adaptors are formed from single stranded oligonucleotides

by annealing complementary single-stranded oligonucleotides that are chemically

synthesized or by enzymatic synthesis. Following synthesis of each strand, the two

oligonucleotide strands are mixed together in a buffered salt solution (*e.g.*, 1 M NaCl,

30   100 mM Tris-HCl pH.8.0, 10 mM EDTA) or in a buffered solution containing Mg^{+2} (*e.g.*,

-10-

10 mM MgCl$_2$) and annealed by heating to high temperature and slow cooling to room
temperature.

The oligonucleotide primer that primes first strand DNA synthesis may comprise a 5'
sequence incapable of hybridizing to a polyA tail of the mRNAs, and a 3' sequence that
5   hybridizes to a portion of the polyA tail of the mRNAs and at least one non-polyA
nucleotide immediately upstream of the polyA tail. The 5' sequence is preferably a
sufficient length that can serve as a primer for amplification. The 5' sequence also
preferably has an average G+C content and does not contain large palindromic sequence;
some palindromes, such as a recognition sequence for a restriction enzyme, may be
10   acceptable. Examples of suitable 5' sequences are CTCTCAAGGATCTACCGCT (SEQ
ID No. _____), CAGGGTAGACGACGCTACGC (SEQ ID No. _____), and
TAATACCGCGCCACATAGCA (SEQ ID No. _____)

The 5' sequence is joined to a 3' sequence comprising sequence that hybridizes to a
portion of the polyA tail of mRNAs and at least one non-polyA nucleotide immediately
15   upstream. Although the polyA-hybridizing sequence is typically a homopolymer of dT
or dU, it need only contain a sufficient number of dT or dU bases to hybridize to polyA
under the conditions employed. Both oligo-dT and oligo-dU primers have been used and
give comparable results. Thus, other bases may be interspersed or concentrated, as long
as hybridization is not impeded. Typically, 12 to 18 bases or 12 to 30 bases of dT or dU
20   will be used. However, as one skilled in the art appreciates, the length need only be
sufficient to obtain hybridization. The non-poly A$^+$ nucleotide is A, C, or G, or a
nucleotide derivative, such as inosinate. If one non-polyA nucleotide is used, then three
oligonucleotide primers are needed to hybridize to all mRNAs. If two non-polyA
nucleotides are used, then 12 primers are needed to hybridize to all mRNAs (AA, AC,
25   AG, AT, CA, CC, CG, CT, GA, GC, GG, GT). If three non-poly A nucleotides are used
then 48 primers are needed (3 X 4 X 4). Although there is no theoretical upper limit on
the number of non-polyA nucleotides, practical considerations make the use of one or
two non-polyA nucleotides preferable.

For cDNA synthesis, the mRNAs are either subdivided into three (if one non-polyA
30   nucleotide is used) or 12 (if two non-polyA nucleotides are used) fractions, each

-11-

containing a single oligonucleotide primer, or the primers may be pooled and contacted with a mRNA preparation. Other subdivisions may alternatively be used. Briefly, first strand cDNA is initiated from the oligonucleotide primer by reverse transcriptase (RTase). As noted above, RASE may be obtained from numerous sources and protocols

5   are well known. Second strand synthesis may be performed by RASE (Gubler and Hoffman, *Gene* 25: 263, 1983), which also has a DNA-directed DNA polymerase activity, with or without a specific primer, by DNA polymerase 1 in conjunction with RNaseH and DNA ligase, or other equivalent methods. The double-stranded cDNA is generally treated by phenol:chloroform extraction and ethanol precipitation to remove

10 protein and free nucleotides.

Double-stranded cDNA is subsequently digested with an agent that cleaves in a sequence-specific manner. Such cleaving agents include restriction enzymes, chemical cleaving agents, triple helix, and any other cleaving agent available. Restriction enzyme digestion is preferred; enzymes that are relatively infrequent cutters (*e.g.*, ≥ 5 bp

15 recognition site) are preferred and those that leave overhanging ends are especially preferred. A restriction enzyme with a six base pair recognition site cuts approximately 8% of cDNAs, so that approximately 12 such restriction enzymes should be needed to digest every cDNA at least once. By using 30 restriction enzymes, digestion of every cDNA is assured.

20 The adapters for use in the present invention are designed such that the two strands are only partially complementary and only one of the nucleic acid strands that the adapter is ligated to can be amplified. Thus, the adapter is partially double-stranded (*i.e.*, comprising two partially hybridized nucleic acid strands), wherein portions of the two strands are non-complementary to each other and portions of the two strands are

25 complementary to each other. Conceptually, the adapter may be "Y-shaped" or "bubble-shaped." When the 5' region is non-paired, the 3' end of other strand cannot be extended by a polymerase to make a complementary copy. The ligated adapter can also be blocked at the 3' end to eliminate extension during subsequent amplifications. Blocking groups include dideoxynucleotides and other available blocking agents. In this type of adapter

30 ("Y-shaped"), the non-complementary portion of the upper strand of the adapters is

-12-

preferably a length that can serve as a primer for amplification. As noted above, the non-complementary portion of the lower strand need only be one base, however, a longer sequence is preferable (*e.g.*, 3 to 20 bases; 3 to 15 bases; 5 to 15 bases, or 14 to 24 bases. The complementary portion of the adapter should be long enough to form a duplex under

5   conditions of ligation.

For "bubble-shaped" adapters, the non-complementary portion of the upper strands is preferably a length that can serve as a primer for amplification. Thus, this portion is preferably 15 to 30 bases. Alternatively, the adapter can have a structure similar to the Y-shaped adapter, but has a 3' end that contains a moiety that a DNA polymerase cannot

10   extend from.

Amplification primers are also used in the present invention. Two different amplification steps are performed in the preferred aspect. In the first, the 3' end (referenced to mRNA) of double stranded cDNA that has been cleaved and ligated with an adapter is amplified. For this amplification, either a single primer or a primer pair is

15   used. The sequence of the single primer comprises at least a portion of the 5' sequence of the oligonucleotide primer used for first strand cDNA synthesis. The portion need only be long enough to serve as an amplification primer. The primer pair consists of a first primer whose sequence comprises at least a portion of the 5' sequence of the oligonucleotide primer as described above; and a second primer whose sequence

20   comprises at least a portion of the sequence of one strand of the adapter in the non-complementary portion. The primer will generally contain all the sequence of the non-complementary potion, but may contain less of the sequence, especially when the non-complementary portion is very long, or more of the sequence, especially when the non-complementary portion is very short. In some embodiments, the primer will contain

25   sequence of the complementary portion, as long as that sequence does not appreciably hybridize to the other strand of the adapter under the amplification conditions employed. For example, in one embodiment, the primer sequence comprises four bases of the complementary region to yield a 19 base primer, and amplification cycles are performed at 56°C (annealing temperature), 72°C (extension temperature), and 94°C (denaturation

30   temperature). In another embodiment, the primer is 25 bases long and has 10 bases of

-13-

sequence in the complementary portion. Amplification cycles for this primer are

performed at 68 °C (annealing and extension temperature) and 94 °C (denaturation

temperature). By using these longer primers, the specificity of priming is increased.

The design of the amplification primers will generally follow well-known guidelines,

5   such as average G-C content, absence of hairpin structures, inability to form primer-

dimers and the like. At times, however, it will be recognized that deviations from such

guidelines may be appropriate or desirable.

In instances where small numbers of cells are available for the initial RNA extraction,

such as small numbers of stem cells, the preferred method of producing a gene expression

10   profile comprises the following general steps. Total RNA is extracted from as few as

5000 stem cells. Using an oligo-dT primer, double stranded cDNA is synthesized and

ligated to an adapter in accordance with the present invention. Using adapter primers, the

cDNA is PCR amplified using the protocol of Baskaran and Weissman (1996) *Genome*

Research 6(7): 633 and/or Liv *et al.* (1992) *Methods of Enzymology.* The original cDNA

15   is therefore amplified several fold so that a large quantity of this cDNA is available for

use in the display protocol according to the present invention. For the display, an aliquot

of this cDNA is incubated with an anchored oligo-dT primer. In one method, this

mixture is first heat denatured and then allowed to remain at 50°C for 5 minutes to allow

the anchor nucleotides of the oligo-dT primers to anneal. This provides for the synthesis

20   of cDNA utilizing Klenow DNA polymerase. The 3'-end region of the parent cDNA

(mainly the polyA region) that remains single stranded due to pairing and subsequent

synthesis of cDNA by the anchored oligo-dT primer at the beginning of the polyA region,

is removed by the 5'-3' exonuclease activity of the T4 DNA polymerase. Following

incubation of the cDNA with T4 DNA polymerase for this purpose, dNTPs are added in

25   the reaction mixture so that the T4 DNA polymerase initiates synthesis of the DNA over

the anchored oligo-dT primer carrying the heel. The net result of this protocol is that the

cDNA with the 3' heel is synthesized for display from the double stranded cDNA as the

starting material, rather than RNA as the starting material as occurs in conventional 3'-

end cDNA display protocol. The cDNA carrying the 3'-end heel is then subjected to

30   restriction enzyme digestion, ligation, and PCR amplification followed by running the

-14-

PCR amplified 3'-end restriction fragments with the Y-shaped adapter on a display gel. An alternate method is presented in Example 1.

After amplification, the lengths of the amplified fragments are determined. Any procedure that separates nucleic acids on the basis of size and allows detection or

5   identification of the nucleic acids is acceptable. Such procedures include slab gel electrophoresis, capillary gel electrophoresis, 2-dimensional electrophoresis, high performance liquid chromatography, and the like.

Electrophoresis is technique based on the mobility of DNA in an electric field. Negatively charged DNA migrates towards a positive electrode at a rate dependent on

10   their total charge, size, and shape. Most often, DNA is electrophoresed in agarose or polyacrylamide gels. For maximal resolution, polyacrylamide is preferred and for maximal linearity, a denaturant, such as urea is present. A typical gel setup uses a 19:1 mixture of acrylamide:bisacrylamide and a Tris-borate buffer. DNA samples are denatured and applied to the gel, which is usually sandwiched between glass plates. A

15   typical procedure can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Approach,* Cold Spring Harbor Press, NY, 1989) or Ausbel *et al.* (*Current Protocols in Molecular Biology,* Greene Publishing Co., NY, 1995). Variations may be substituted as long as sufficient resolution is obtained.

Capillary electrophoresis (CE) in its various manifestations (free solution,

20   isotachophoresis, isoelectric focusing, polyacrylamide get. micellar electrokinetic "chromatography") allows high resolution separation of very small sample volumes. Briefly, in capillary electrophoresis, a neutral coated capillary, such as a 50 μm X 37 cm column (eCAP neutral, Beckman Instruments, CA), is filled with a linear polyacrylamide (*e.g.,* 0.2% polyacrylamide), a sample is introduced by high-pressure injection followed

25   by an injection of running buffer (*e.g.,* 1X TBE). The sample is electrophoresed and fragments are detected. An order of magnitude increase can be achieved with the use of capillary electrophoresis. Capillaries may be used in parallel for increased throughput (Smith et al. (1990) *Nuc. Acids. Res.* 18:4417; Mathies and Huang (1992) *Nature* 359:167). Because of the small sample volume that can be loaded onto a capillary,

30   sample may be concentrated to increase level of detection. One means of concentration

-15-

is sample stacking (Chien and Burgi (1992) *Anal. Chem* 64:489A). In sample stacking, a large volume of sample in a low concentration buffer is introduced to the capillary column. The capillary is then filled with a buffer of the same composition, but at higher concentration, such that when the sample ions reach the capillary buffer with a lower

5   electric field, they stack into a concentrated zone. Sample stacking can increase detection by one to three orders of magnitude. Other methods of concentration, such as isotachophoresis, may also be used.

High-performance liquid chromatography (HPLC) is a chromatographic separation technique that separates compounds in solution. HPLC instruments consist of a reservoir

10   of mobile phase, a pump, an injector, a separation column, and a detector. Compounds are separated by injecting an aliquot of the sample mixture onto the column. The different components in the mixture pass through the column at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase. IP-RO-HPLC on non-porous PS/DVB particles with chemically

15   bonded alkyl chains can also be used to analyze nucleic acid molecules on the basis of size (Huber et al. (1993) *Anal. Biochem.* 121:351; Huber et al. (1993) *Nuc. Acids Res.* 21:1061; Huber et al. (1993) *Biotechniques* 16:898).

In each of these analysis techniques, the amplified fragments are detected. A variety of labels can be used to assist in detection. Such labels include, but are not limited to,

20   radioactive molecules (*e.g.*, ^{35}S, ^{32}P, ^{33}P), fluorescent molecules, and mass spectrometric tags. The labels may be attached to the oligonucleotide primers or to nucleotides that are incorporated during DNA synthesis, including amplification.

Radioactive nucleotides may be obtained from commercial sources; radioactive primers may be readily generated by transfer of label from γ-^{32}P-ATP to a 5'-OH group

25   by a kinase (*e.g.*, T4 polynucleotide kinase). Detection systems include autoradiograph, phosphor image analysis and the like.

Fluorescent nucleotides may be obtained from commercial sources (*e.g.*, ABI, Foster city, CA) or generated by chemical reaction using appropriately derivatized dyes. Oligonucleotide primers can be labeled, for example, using succinimidyl esters to

30   conjugate to amine-modified oligonucleotides. A variety of florescent dyes may be used,

-16-

including 6 carboxyfluorescein, other carboxyfluorescein derivatives, carboxyrhodamine

derivatives, Texas red derivatives, and the like. Detection systems include

photomultiplier tubes with appropriate wave-length filters for the dyes used. DNA

sequence analysis systems, such as produced by ABI (Foster City, CA), may be used.

5   After separation of the amplified cDNA fragments, cDNA fragments which

correspond to differentially expressed mRNA species are isolated, reamplified and

sequenced according to standard procedures. For instance, bands corresponding the

cDNA fragments can be cut from the electrophoresis gel, reamplified and subcloned into

any available vector, including pCRscript using the PCR script cloning kit (Stratagene).

10   The insert is then sequenced using standard procedures, such as cycle sequencing on an

ABI sequencer (Foster City, CA).

An additional means of analysis comprises hybridization of the amplified fragments to

one or more sets of oligonucleotides immobilized on a solid substrate. Historically, the

solid substrate is a membrane, such as nitrocellulose or nylon. More recently, the

15   substrate is a silicon wafer or a borosilicate slide. The substrate may be porous (Beattie

et al. WO 95/11755) or solid. Oligonucleotides are synthesized in situ or synthesized

prior to deposition on the substrate using standard procedures. Various chemistries are

known for attaching oligonucleotides. Many of these attachment chemistries rely upon

functionalizing oligonucleotides to contain a primary amine group. The oligonucleotides

20   are arranged in an array form, such that the position of each oligonucleotide sequence can

be determined.

The amplified fragments, which are generally labeled according to one of the methods

described herein, are denatured and applied to the oligonucleotides on the substrate under

appropriate salt and temperature conditions. In certain embodiments, the conditions are

25   chosen to favor hybridization of exact complementary matches and disfavor hybridization

of mismatches. Unhybridized nucleic acids are washed off and the hybridized molecules

detected, generally both for position and quantity. The detection method will depend

upon the label used. Radioactive labels, fluorescent labels and mass spectrometry label

are among the suitable labels.

-17-

The present invention as set forth in the specific embodiments, includes methods to identify a therapeutic agent that modulates the expression of at least one stem cell gene associated with the differentiation, proliferation and/or survival of stem cells.

As an example, the method to identify an agent that modulates the expression of at

5   least one stem cell gene associated with the differentiation of a stem cell population, comprises the steps of preparing a first gene expression profile of an undifferentiated stem cell population, preparing a second gene expression profile of a stem cell population at a defined stage of differentiation, treating said undifferentiated stem cell population with the agent, preparing a third gene expression profile of the treated stem cell

10   population, and comparing the first, second and third gene expression profiles. Comparison of the three gene expression profiles for RNA species as represented by cDNA fragments that are differentially expressed upon addition of the agent to the undifferentiated stem cell population identifies agents that modulate the expression of a least one gene in undifferentiated stem cells that is associated with stem cell

15   differentiation.

While the above methods for identifying a therapeutic agent comprise the comparison of gene expression profiles from treated and not-treated stem cells, many other variations are immediately envisioned by one of ordinary skill in the art. As an example, as a variation of a method to identify a therapeutic agent that modulates the expression of at

20   least one stem cell gene associated with the differentiation, the second gene expression profile of a stem cell population at a defined stage of differentiation and the third gene expression profile of the treated stem cell population can each be independently normalized using the first gene expression profile prepared from the undifferentiated stem cell population. Normalization of the profiles can easily be achieved by scanning

25   autoradiographs corresponding to each profile, and subtracting the digitized values corresponding to each band on the autoradiograph from undifferentiated stem cells from the digitized value for each corresponding band on autoradiographs corresponding to the second and third gene expression profiles. After normalization, the second and third gene expression profiles can be compared directly to detect cDNA fragments which

-18-

correspond to mRNA species which are specifically expressed during differentiation of a stem cell population.

Specific Embodiments

Example 1

5   *Production of gene expression profiles generated from cDNAs made with RNA isolated from undifferentiated and partially differentiated stem cells.*

Crude Marrow Preparation

Expression profiles of RNA expression levels from undifferentiated stem cells and stems cells at various levels of differentiation, including partially differentiated and

10   terminally differentiated stem cells, offer a powerful means of identifying genes whose expression levels are associated with stem cell differentiation or proliferation. As an example, the production of expression profiles from murine lineage negative, rhodamine low, Hoechst low and rhodamine bright, Hoechst low hematopoietic precursor cells allows for the identification of mRNA species and their encoding genes whose

15   expression levels are associated with stem cell differentiation

Hoechst^{low}/Rhodamine^{low} hematopoietic stem cells were isolated by sacrificing 30 Balb/c female mice (6-12 weeks) and surgically removing the iliac crests, femurs and tibiae. The bones were cleaned and placed in 10 ml PBS/5% HI-FBS on ice. One tube was used for the bones from 10 mice. The bones were ground throughly with a pestle

20   until completely broken. Following grinding, the supernatant was removed into a 50 ml conical tube through a 40 μM filer(Falcon #2340). 10 ml PBS/FBS was added to the mix and the supernatant removed. The supernatant was then centrifuged (1250 rpm) for 5-10 minutes. The supernatant which contains a high concentration of lipid was then decanted and discarded.

25   The cells were then pooled into 25 or 50 ml fresh PBS/FBS, and tiny bone fragments removed by settling. The cells were then counted in crystal violet. Cells were diluted and underlayed with LSM, centrifuged at 2000rpm(1000xg) for 20 minutes. To harvest the buffy coat, the supernatant was removed to within 1 cm of the cells. The next 8-

-19-

10ml of medium and cells were harvested by swirling the media around in the tube to
draw cells from all sides of the gradient. The cell volume was then brought up to 50 ml
with PBS/FBS and spun at 1400rpm 5-10 minutes.

Lineage Depletion

5   Cells were counted in Crystal Violet and resuspended in fresh PBS/FBS. Lineage-
specific antibodies were added as follows:

TER 119	0.1µg/ml final concentration
B220	15µl/10^8 cells
Mac-1	15µl/10^8 cells.
Gr-1	15µl/10^8 cells
Lyt-2	1/20 final dilution
L3T4	1/20 final dilution
Yw25.12.7	1/100 final dilution

The cells were incubated on ice for 15 minutes, brought to a volume of 50ml with

15   PBS/FBS and collected at 1400rpm for 5-10 minutes, and washed to remove unbound
antibodies.

During the antibody binding step, Magnetic Beads(Dynabeads M-450) were prepared at a
ratio of 5 beads/cell. The beads were coated with Sheep anti-Rat antibodies that bind to
the lineage-specific antibodies, which are all of rat origin. When the beads are placed in

20   a magnetic field, the Lin$^+$ cells are removed. The resulting supernatant contains the Lin$^-$
population (granulocytes and lymphocyte populations will be substantially depleted or
absent after this step.)

Hoechst/Rhodamine Staining

Rhodamine 123 was added to a final concentration of 0.1 µg/ml, then incubated at

25   32°C for 20 minutes in the dark. Without further manipulation or washing, HOECHST
33342 was added to a final concentration of 10µM then incubated at 37°C for an
additional hour. The aliquot of crude marrow was brought to 0.5 ml with PBS/FBS and
Hoechst to this cell preparation as well. The volume was brought to 50 ml with
PBS/FBS, centrifuged at 1400rpm for 5-10 minutes, supernatant discarded and cells

30   resuspended to 2x10^7 cells/ml.   The rhodamine only and Hoechst Only/Crude Marrow

-20-

were washed in parallel. These two populations were then resuspended in 0.5ml

PBS/FBS for flow cytometry analysis

Total RNA was extracted from approximately 5000 stem cells. Using an oligo-dT

primer, double stranded cDNA is synthesized and ligated to an adapter in accordance

5   with the present invention. Using adapter primers, the cDNA is PCR amplified using the

protocol of Baskaran and Weissman (1996) *Genome Research* 6(7): 633 and Lie *et al.*,

Methods of Enzymology, ____. The original cDNA is therefore amplified several fold so

that a large quantity of this cDNA is available for use in the display protocol according to

the present invention.

10   Synthesis of cDNA for the gene expression profiles was performed as below:

Materials and Reagents

A microPoly(A)Pure mRNA Isolation kit (Ambion Inc.) was used for mRNA isolation.

All the reagents for cDNA synthesis were obtained from Life Technologies Inc. Klentaq1

DNA polymerase (25U/μl) was from Ab peptides Inc. Native *Pfu* DNA polymerase

15   (2.5U/μl) was purchased from Stratagene Inc. Betaine monohydrate was from Fluka

BioChemica and dimethylsulfoxide (DMSO) was from Sigma Chemical Company.

Deoxynucleoside triphophates (dNTPs, 100mM) and bovine serum albumin (BSA, 10

mg/ml) were purchased from New England Biolabs, Inc. Qiaquick PCR purification kit

(Qiagen) was used to purify the amplified PCR products. The oligonucleotides used in the

20   Examples were synthesized and gel purified in the DNA synthesis laboratory (Department

of Pathology, Yale University School of Medicine, New Haven, CT).

Table 1. Sequences of oligonucleotides.

T_7-SalI-oligo-d(T)V	5'-ACG TAA TAC GAC TCA CTA TAG GGC GAA TTG GGT CGA C-d (T) $_{18}$ V-3' , where V = A, C, G
anti-NotI Long	5'-CTT ACA GCG GCC GCT TGG ACG-3'

-21-

NotI Short	5'-AGC GGC CGC TGT AAG-3'
NotI/RI primer	5'-GCG GAA TTC CGT CCA AGC GGC CGC TGT AAG-3'

Methods

I. Preparation of mRNA

5   MicroPoly(A)Pure mRNA isolation kit was used for the isolation of Poly(A)$^+$ RNA
following the kit instructions. mRNA from a small number of mouse hematopoietic cells
(5,000-10,000 cells) was extracted, eluted from the column, and precipitated by adding 0.1
volume of 5M ammonium acetate and 2.5 volumes of chilled ethanol with 2μg glycogen as
carrier. The tubes were left at -20°C overnight. The pellets were collected by centrifugation
10   at top speed for 30 minutes, washed with 70% ethanol and air-dried at room temperature.
The pellets were resuspended in 10μl H_2O/0.1mM EDTA solution. We observed that the
dissolved mRNA solution was cloudy due to the leaching of column materials, therefore the
samples were centrifuged at 4°C for 5 minutes. The supernatant was collected for further
use.

15   II. cDNA synthesis
        First strand cDNA synthesis
        The cDNA synthesis reaction (final reaction volume is 20μl) was carried out as
described in the instruction manual (Superscript Choice System) provided by Life
Technologies Inc. For the first strand cDNA synthesis, mRNA (10μl) isolated from a small
20   number of cells was annealed with 200ng (1μl) of T_7-SalI-oligo-d(T)V-primer (see Table-1)
in a 0.5-ml micro centrifuge tube (no stick, USA Scientific Plastics) by heating the tubes at
65°C for 5 minutes, followed by quick chilling on ice for 5 minutes. This step was repeated

-22-

once and the contents were collected at the bottom of the tube by a brief centrifugation. The

following components were added to the primer annealed mRNA on ice prior to initiating

the reaction, 1μl of 10mM dNTPs, 4μl of 5 x first strand buffer [250mM Tris-HCl (pH 8.3),

375mM KCl, 15mM MgCl$_2$], 2μl of 100mM DTT and 1μl of RNase Inhibitor (40U/μl). All

5   the contents were mixed gently and the tubes were pre-warmed at 45⁰C for 2 minutes. The

cDNA synthesis was initiated by adding 200 units (1μl) of Superscript II Reverse

Transcriptase and the incubation continued at 45°C for 1 hour.

Second strand cDNA synthesis

At the end of first strand cDNA synthesis, the tubes were kept on ice. Second

10   strand cDNA synthesis reaction (final volume is 150μl) was set up in the same tube on

ice by adding 91μl of nuclease free water, 30μl of 5x second strand buffer [100mM

Tris-HCl (pH 6.9), 23mM MgCl$_2$, 450mM KCl, 0.75mM (β-NAD$^+$ and 50mM

ammonium sulfate], 3μl of 10mM dNTPs, 1μl of *E.coli* DNA ligase (10U/μl), 4μl of

E.coli DNA polymerase I (10U/μl) and 1μl of *E.coli* RNase H (2U/μl). The contents were

15   mixed gently and the tubes were incubated at 16°C for 2 hours. Following the incubation,

the tubes were kept on ice, 2μl of T$_4$ DNA polymerase (3U/μl) was added and the

incubation was continued for another 5 minutes at 16°C. The reaction was stopped by the

addition of 10μl of 0.5M EDTA (pH 8.0) and extracted once with equal volume of

phenol: chloroform 1:1 (v/v) and once with chloroform. The aqueous phase was then

20   transferred to a new tube and precipitated by adding 0.5 volumes of 7.5M ammonium

acetate (pH 7.6), 2μg of glycogen (as carrier) and 2.5 volumes of chilled ethanol. The

samples were left at -20⁰C for overnight and the cDNA pellets were collected by

centrifugation at top speed for 20 minutes. The pellets were washed once with 70%

ethanol, air-dried and dissolved in 14μl of nuclease free water.

25   As the amount of cDNA derived from a small number of cells may be low, it may

be necessary to amplify the cDNA for further analysis. To uniformly amplify the cDNA,

an adaptor (NotI adaptor) was first ligated to both ends of the cDNA. Following adaptor

-23-

ligation, the cDNAs were amplified with NotI/RI primer (see *table 1*), by a modified

PCR method using betaine and DMSO.

Ligation of cDNA with NotI adaptor

Preparation of NotI adaptor: The NotI adaptor was prepared by annealing

5   NotI-short and anti-NotI-long oligonucleotides (see Table 1). The anti-NotI-long

oligonucleotide was phosphorylated to ensure that both the adaptor oligonucleotides are

ligated to the cDNA. $1\mu g$ of anti-NotI-long was mixed with $1\mu l$ of 10x T_4 polynucleotide

kinase buffer [700mM Tris-HCl (pH 7.6), 100mM $MgCl_2$ and 50mM DTT], $1\mu l$ of

10mM adenosine triphosphate (ATP), adjusted the volume to $9\mu l$ with water and the

10   reaction was initiated by adding $1\mu l$ of T_4 polynucleotide kinase ($10U/\mu l$). The tubes were

incubated at 37°C for 30 minutes and then the enzyme was inactivated at 65°C for 20

minutes. The annealing was carried out by adding the following components to the above

phosphorylated anti-NotI-long: $1\mu g$ of NotI-short, $2\mu l$ of 10x oligo annealing buffer

[100mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0) and 1M NaCl] and water to adjust

15   the final volume to $20\mu l$. The sample was heated at 65°C for 10 minutes and allowed to

cool down to room temperature. The annealed adaptor was stored at -20°C.

Ligation of cDNA with annealed NotI adaptor: To set up this reaction,

$14\mu l$ of cDNA was mixed with 100ng of annealed NotI adaptor in a 0.5-ml micro

centrifuge tube. To this mixture $2\mu l$ of 10x T_4 DNA ligase buffer [500mM Tris-HCl (pH

20   7.8), 100mM $MgCl_2$, 100mM DDT, 10mM ATP and 250mg/ml BSA] was added and

adjusted the volume with water to $18\mu l$ and mixed gently. The reaction was initiated by

adding $2\mu l$ of T_4 DNA ligase ($400U/\mu l$) and incubated at 16°C overnight.

III.   cDNA amplification

A modified betaine-DMSO PCR method (Baskaran *et al.* (1996)) Genome

25   Research 6:633) was used to uniformly amplify the cDNA with different GC content.

This method uses the LA system, which combines a highly thermostable form of *Taq*

DNA polymerase (Klentaq1, which is devoid of 5'-exonuclease activity) and a

proofreading enzyme (*Pfu* DNA polymerase, which has 3'-exonuclease activity). The

-24-

. LA16 enzyme consists of 1 part of *Pfu* DNA polymerase and 15 parts of KlenTaq1 DNA

Polymerase (v/v). The NotI adaptor-ligated cDNA was diluted 10 fold with water. 2 μl of

this diluted cDNA was used as the template for PCR. The PCR reaction (50μl final

volume) was set up with the following components: 5μl of 10x PCR buffer [200mM

5   Tris-HCl (pH 9.0), 160mM ammonium sulfate and 25mM MgCl$_2$], 16μl of water, 0.8μl

of BSA (10mg/ml), 1μl of NotI/RI PCR primer (100ng/ul), 5μl of 50% DMSO (v/v), 15μl

of 5M Betaine and 0.2μl of LA16 enzyme. These components were mixed gently on ice

and then heated to 95°C for 15 seconds on a PCR machine, and held at 80°C while 5μl of

2mM dNTPs were added to start the reaction. The PCR conditions were as follows: *Stage*

10   *1:* 95°C for 15 seconds, 55°C for 1 minute, 68°C for 5 minutes, 5 cycles. *Stage 2:* 95°C

for 15 seconds, 60°C for 1 minute, 68°C for 5 minutes, 15 cycles.

         After amplification, cDNA was purified with the Qiaquick PCR purification kit

(following the instructions provided by the supplier). The purified cDNA was eluted in

the desired volume of water.

15   Gene expression profiles were prepared from the purified cDNA as previously

described by Prashar *et al.* in WO 97/05286 and in Prashar *et al.* (1996) *Proc. Natl. Acad.*

Sci. USA 93:659-663. Briefly, the adapter oligonucleotide sequences were

CTTACAGCGGCCGCTTGGACG, GAATGTCGCCGGCGA or alternatively,

A1 (TAGCGTCCGGCGCAGCGACGGCCAG) and

20   A2 (GATCCTGGCCGTCGGCTGTCTGTCGGCGC). When A1/A2 were used, one

microgram of oligonucleotide A2 was first phosphorylated at the 5' end using T4

polynucleotide kinase (PNK). After phosphorylation, PNK was heated denatured, and

1μg of the oligonucleotide A1 was added along with 10× annealing buffer (1 M

NaCl/100 mM Tris-HCl, pH8.0/10 mM EDTA, pH8.0) in a final vol of 20 μl. This

25   mixture was then heated at 65°C for 10 min followed by slow cooling to room

temperature for 30 min, resulting in formation of the Y adapter at a final concentration of

100 ng/μl. About 20 ng of the cDNA was digested with 4 units of a restriction enzyme

such as *Cla*I, *Bgl* II, etc. in a final vol of 10 μl for 30 min at 37°C. Two microliters ($\approx$4

ng of digested cDNA) of this reaction mixture was then used for ligation to 100 ng ($\approx$50-

30   fold) of the Y-shaped adapter in a final vol of 5μl for 16 hr at 15°C. After ligation, the

-25-

reaction mixture was diluted with water to a final vol of 80 μl (adapter ligated cDNA

concentration, $\approx$50 pg/μl) and heated at 65°C for 10 min to denature T4 DNA ligase, and

2-μl aliquots (with $\approx$100 pg of cDNA) were used for PCR.

The following sets of primers were used for PCR amplification of the adapter

5   ligated 3'-end cDNAs: GCGGAATTCCGTCCAAGCGGCCGCTGTAAG or

alternatively, RP 5.0 (CTCTCAAGGATCTTACCGCTT $_{18}$AT), RP 6.0

(TAATACCGCGCCACATAGCAT $_{18}$CG), or RP 9.2

(CAGGGTAGACGACGCTACGCT$_{18}$GA) were used as 3' primer while A1.1

(TAGCGTCCGGCGCAGCGAC) served as the 5' primer. To detect the PCR products

10   on the display gel, 24 pmol of oligonucleotide A1.1 was 5'-end-labeled using 15 μl of

[γ-^{32}P]ATP (Amersham; 3000 Ci/mmol) and PNK in a final volume of 20 μl for 30 min

at 37°C. After heat denaturing PNK at 65°C for 20 min, the labeled oligonucleotide was

diluted to a final concentration of 2 μM in 80 μl with unlabeled oligonucleotide A1.1.

The PCR mixture (20μl) consisted of 2 μl ($\approx$100 pg) of the template, 2μl of 10$\times$ PCR

15   buffer (100 mM Tris·HCl, pH 8.3/500 mM KCl), 2 μl of 15 mM MgCl$_2$ to yield 1.5 mM

final Mg^{2+} concentration optimum in the reaction mixture, 200 μM dNTPs, 200 nM each

5' and 3' PCR primers, and 1 unit of Amplitaq. Primers and dNTPs were added after

preheating the reaction mixture containing the rest of the components at 85°C. This "hot

start" PCR was done to avoid artefactual amplification arising out of arbitrary annealing

20   of PCR primers at lower temperature during transition from room temperature to 94°C in

the first PCR cycle. PCR consisted of 28-30 cycles of 94°C for 30 sec, 50°C for 2 min,

and 72°C for 30 sec. A higher number of cycles resulted in smeary gel patterns. PCR

products (2.5μl) were analyzed on 6% polyacrylamide sequencing gel. For double or

multiple digestion following adapter ligation, 13.2 μl of the ligated cDNA sample was

25   digested with a secondary restriction enzyme(s) in a final vol of 20 μl. From this

solution, 3μl was used as template for PCR. This template vol of 3 μl carried $\approx$ 100 pg

of the cDNA and 10 mM MgCl$_2$ (from the 10$\times$ enzyme buffer), which diluted to the

optimum of 1.5 mM in the final PCR vol of 20 μl. Since Mg^{2+} comes from the

restriction enzyme buffer, it was not included in the reaction mixture when amplifying

30   secondarily cut cDNA. Bands may then be extracted from the display gels as described

-26-

by Liang *et al.* (1995 *Curr. Opin. Immunol.* 7:274-280), reamplified using the 5' and 3'

primers, and subcloned into pCR-Script with high efficiency using the PCR-Script

cloning kit from Stratagene. Plasmids were sequenced by cycle sequencing on an ABI

automated sequencer.

5   Figure 1 presents an autoradiogram of the gene expression profiles generated

from cDNAs made with RNA isolated from Lin^+, LRH, LRH48 and LRBRH cells. All

possible 12 anchoring oligo $d(T)n1$, $n2$ were used to generate a complete expression

profile for the enzyme *ClaI*.

Table 2 presents the sequences of numerous differentially expressed bands from

10   expression profiles made from LIN^+, LRH, LRH48 and LRBRH.

TABLE 2

HSC-DD-006	TTTAATTAGCGCTCTATATACATTGCG GAACTTCCCCCGACTGCAGCAGTTTGA CTTTGGCACAACATCAAGTTCCATTTC TTTTGGACATTGGATTCTGTTTTGANA GTATGTATGCCCCAAAGCATTTTCAGT GTCATCAGGATTAGTTGGGCCCATTCA CAGTAATTCANANATC
HSC-DD-285	TAGAATACCTGGATGGCTTCTCTTGTC CACCCGATCTCCCGTGTTACCAATGTG TATGGTCTCCTTCTCCCGAAAGTGTAC TTAATCTTTGCTTTCTTTGCACAATGTC TTTGGTTGCAAGTCATAAGCCTGAGGC AAATAAAATTCC

-27-

HSC-DD-007B	GATCTGGCTAGACAGTTATTCTGAACT ATGGCTTCAAGATGAACAAGACAAGC CTAAAAGGATGGAGAGAGGCAATGGA GATAATGTTTTGGAGGAAGTATGTCAC TCAAGCATGAACTCTGTTTATTTAGAA ATGAGATTCCATATATGTGGTACATGT GGAAAGAATCTAAAAAGTCCTTTAAA TTTTTTCATTCCAAAAG
HSC-DD-238	CTNNANNAGCACTCTTCTTGGCCAGAC CTCTGTCCAAGGCTCATTAGAAAGCTG GGGTTNTGTNCACGTNACNNACTTNAT CNAAACTNTTGCTGTNTTGGCATAAGT TGTGTNTCTGGACTGTNNTGTATTCCC CTCTAGACAAAGGANCAACNNAAAAG TNNTTGCNNNCTTTNCCAGAACATNCT CAAAGCCTNTGATGGAGGAGCACAAG GACCCTGTCTGCTGAGGGCCCATGGNT CCTCTCAGGGGTTTCTNCCCACCNAGG CAGTGCCTTCATTNGCTAGTNGTNCAG TTACTTGTAGNTTATCTTTNAATAAAT TTNAATAAAANCTA
HSC-DD-206	CTAGATTGTGTGGTTTGCCTCATTGTG CTATTTGCGCACTTTCCTTCCCTGAAG AAATANCTGTGAANCTTCTTTCTGTTC AGTCCTAANATTCNAAATANAGTGAG ACTATG

-28-

HSC-DD-214	CTCAAGNACGGGCCAGGTAAGGGCCT TTAACACAACTAAATCAAGGTGTGCTT NCCTCCGGGTTCTATGCAAGCAAGGCA TACACACTGCACTCTCNCNCTCNCTAA ACTGGAAANGTACAGTNGCAGGGCTG GTTTCAGACNACGTGATGCNTGTTTAC AAAC
HSC-DD-035	TTTTTATTCAATATATTAAATATATTAA TCAGAAAAGTCACATCCTATAAATCCA GGAAAATACACAAATATAAATCAGAA TCTGTCAATCACCTTCTTGAGTGACAG TTATGTACACATGGAAGGAGAGCGGA AGAGATC
HSC-DD-129	CGATATACACCATCGGTCTGGGGCCAA CGCTAATACTACTTGGTGCTGCCAATT GAATTCTGGTTTGCTGTGAATCTCTAT CAACAAGAGTATCATTTGTGAATGCTT TAATTTATTGAGAAAGAACAAGAAGA TGATGGATACATTGATACATTTGCGCA GCCTTGCAGCCTGACTCAATTCTGCTG TTCATCAGTTTTAATGTCCTTTCTGTGT CATACGTG

HSC-DD-040	GATCTTTTTCCTTCACTTATTGCTGAA ACCAAGNGCACAATTCCCATTAAGNG AAGGATCTCTGTGCTGTAAACTAAACA AATTGTGCATTTTTCTGGGGCCATTG TTTTTGGTTTATTTTGTTATTTTGTTTTG TTTTTGTTTTTTTGGTTTCATTTTGTTTT GGGTTGGTCCAATTTTAAAAGGAAATA CTACAATAAAAATGTTA
HSC-DD-011	GATCTGATTTGCTAGTTCTTCCTGGTA GAGTTATAAATGGAAAGATTACACTAT CTGATTAATAGTTTCTTCATACTCTGC ATATAATTTGTGGCTGCAGAATATTGT AATTTGTTGCACACTATGTAACAAAAC TGAAGATATGTTTAATAAATATTGTAC T
HSC-DD-121	GCGATGTTCTTCTACTCACAACTCACG TTGGTGGCCTGGGCCTGAACTTGACTG GAGCTGACACTGTGGTGTTTGTGGAGC ATGACTGGAACCCTATGCGAGATCTGC AGGCCATGGACCGGGCCCATCGTATTG GGCAGAAACGTGTGGTTAATGTCTACC GGTTGATAACCAGA
HSC-DD-015B	GATCTGGAAGGGAATGTCCAAAGAGA AGAAGGAGGAGTGGGACCGCAAGGCT GAGGATGCTAGGAGGGAGTATGAGAA AGCCATGAAAGAGTATGAAGGAGGAA GAGGGGACTCATCTAAAAG

-30-

HSC-DD-039	GATCTTCGACACAGAGAAGGAGAAAT ACGAGATTACAGAGCAGCGAAAGGCT GACCAGAAAGCTGTGGATTTGCAGATT TTGCCAAAGATTAAAGCTGTTCCTCAG CTCCAGGGCTACCTGCGCTCTCAGTTT TCCCTGACAAACGGGATGTATCCTCAC AAACTGGTCTTCTAAATTGTTAACCTA ATTAAACAG
HSC-DD-042	ACTCAATCTCTTCAAACTCTTTATACT GGNCTATNATNAGNGGGGATGTGNCA ANATNGACNCTGGTGGTGTATGAAAG AAAAGNTCNATGGACNTNGGCATNCC AAGATTGAATTCACCTGCTTCCTACGA TGTGTGAAACTGCTAATAGCAAAATAT CTCTANGGTTATGANGAGTACTGTCGT TCTGCAAATATTCACTTCANAACTANN CACCACGTTNAA
HSC-DD-256A	CTAGATAATCCCTTACTGAGTCTTTCTT CNCAGGTGATTCANTTGAGTTGACAAT TANNNCTAAGAATTCAATGGACTANT GAGGTGCCTCAGCAGNTAATAGCANT TGCTGTTCTTCCAGAGGACCAGAGTTC AGTTTCTCATCCCAAGTTGGGCTGCTC GTNAGTGTCGGTAANTCCAGCTTCAGG GGCTTGAATTTATACTGACCATGGGCA CCTGTACCCCAACACANACACATACA CAT

-31-

HSC-DD-256B	CTAGAAGTTAATCCTGTNAAGCATGGT AAGAATANCATTCTCAANATCTTGAGT TAANAAAGATCTTGGAGGNGGCTGGN GAGATGGCTCANTGGTTAAGANCNCT GACTGCTCTTCCAGAGGTCCTGANTTC AATTCCCANCAACCACATGGTGGNTCA CAACCANCTGTAATGATACCTGATGCC ATCNTCCGTGGTGTATCTGAANACANC TACAGTGACAGCTACANCG
HSC-DD-045	GGATTTTATTCTAGGCTTGGCCAGATA CAGGTTGGCATCCTAGGGGAGGAAGA TAACAATGTCATAGGTGAATTTGTTAG GAGAGGCAAGACATGGGAAATCATTG ATTTCTTCAGATTTCTTTAAAGCAAAT TAGAAGATAAATGTCTAAAAGAGATA CACTTAAAAAATGGTGAAACTATAAC CCCTTAAGGAGAGCCAGATGTGGCAG GAGCCAGGTCTGAAAATGGTAGCTGA AGTAAGCAGACCAGCGTAAGATC
HSC-DD-068	CGATGAGTCAGAGAGGAAGTGGACAG TGCGTTATTCATTACAGCAAAGGATTT CGTTGGCATCAAAATCTAAGTTTGTTT TACAAAGATTGTTTTTAGTACTAAGCT GCCTTGGCAGTTTGCATTTTTGAGCCA AACAAAAATATATTATTTTC

-32-

HSC-DD-143	CGATTCAATTGTATAAATGATTATAAT TTCTTTCATGGAAGCATGATCCTTCTG ATTAAGAACTGTACCCCATATTTTATG CTGGTTGTCTGCAAGCTTGTGCGATGA TGTTATGTTCATGTTAATCCTATTTGTA AAATGAAGTGTTCCTGACCTTATGTTA AAAAGAGAGAAGTAAATAACAGACAT TATTCAGTTATTTTGTCCTTTATCGAAA AACCAGATTTCATTTTTCCTTTTTGTTT GTGATCTCATTTGGAAATAATTGGCAA GTTGAGGTACTTTCTTCCCATGCTTTGT ACAATATAAACTGTTATGCCTTTCAGT GCGTTACTGTGGG
HSC-DD-263A	CTAGAGGTGGGAACTGGCTCCACTCCA CACAGCAGCCAGTTAGTTAGTGACGGT CAGCTGCATGCAGGGGAATGAAGGAC TCGGAGAGAACGTTCTGTGCTATGTGT GTTCCATAGAGATTAAAAAGGAGGCC TGGAGCCGAGCATGGTGGTGCACGCC TTTAATCCCAGCACTTGGGAGGCAGAG TCAGGTGGATTTCTGAGTTCATTGCCA GCCTGGTCTACAGAGTGAATTCCAGGA CAGGCAGGGCTACACAGAGAAACCCT GTCTCAAAAAA
HSC-DD-263B	CTAGAATTTGCAGTAGCATTAATTCAA GCCTACGTATTCACCCTCCTAGTAAGC CTATATCTACAT

-33-

HSC-DD-239A1	CTAGACATAAGATATTGTACATAAAG ANAATTTTTTTTGCCTTTAAATAGATA AAAGTATCTATCAGATAAAAATCANG TTGTAAGTTATATTGAAGACAATTTGA TACATAATAAAAGAT
HSC-DD-239A1'	GGGGAGNNNNCNAGNAANNAGANTC GTACGTAAANAGAANNNTGGTGCNTT TANATAGAAAANGTACTATCANATAA NAATCAGGTTGTAAGTTATATTGAAGA CGNTTTGATACATAATAAAAGAT
HSC-DD-261	CTAGACTGACAAAGACTTTTTGTCAAC TTGTACAATCTGAAGCAATGTCTGGCC CACAGACAGCTGAGCTGTAAACAAAT GTCACATGGAAATAAATACTTTATC
HSC-DD-028A	CTCTCTTGCCACCCAGATGGTTAGGAT GATTCTGAAGATGATGACATCCGTAAG CCTGGAGAATCTGAAGAATAAACTGT ACCAT
HSC-DD-021	ATCTCTGGCAGGTCAAGTCTGGGACAA TCTTTGACAATTTCCTCATCACCAGTG ATGAGGCCTATGCAGCCAGTTCTAGCG CAGCTCACACTGAGAGTGTAAGAACT ACGAACAAAATNTCTATTAAATTAAG

5

HSC-DD-025	GATCTCGGAATGGACCCAACTGCTCCT GCTCCACCGGCGGCTCCTGCACTTGCA CCAGCTCCTGCGCCTGCAAGAACTGCA AGTGCACCTCCTGCAAGAAGAGCTGCT GCTCCTGCTGTCCCGTGGGCTGCTCCA AATGTGCCCAGGGCTGTGTCTGCAAAG GCGCCGCGGACAAGTGCACGTGCTGT GCCTGATGTGACGAACAGCGCTGCCA CCACGTGTAAATAGTATCGGACCAACC CAGCGTCTTCCTATACAGTTCCACCCT GTTTACTAAACCCCCGTTTTCTACCGA GTACGTGAATAATAAAAGCCT
HSC-DD-077	ATTCAGACGAATGAGACTCCTCCACAT TGGAGACAAGAGATGCAGAGAGCTCA GAGAATGAGGGTGTCAAGTGGTGAAA GATGGATCAAAGGGGATAAGAGTGAG TTAAATGAAATAAAAGAAAATCAAAG GAGCC
HSC-DD-245	NGCNNNNNNNCCAGNAGGAGGAGAA GATGACTGGCCAGTATCANAATGGGA TAAGATGAGGCGCGCCCTGGAGTACA CCATCTACAACCAGGAGCTCAACGAG ACGCGCGCTAAGCTCGACGAGCTTTCT GCTAANCGAGAAACNAGTGGAGAGAA ATCCNGACAACTAAGGGATGCCCAGC AGGATGCANGAGACAAAATGGAGGAT ATTGAGCGCCAGGTTAGAGAACTGAA AACAATNAT

-35-

HSC-DD-226	CTCAAGGAAAAGACAGCACCNCGTGC CTGGCATCTGNTGNNTTAGNTNATNTN NAANTNTCNNNTNGNCCTGGCAACGG TTCCTGAACNAATTACCACTCCTTCTT GCCAGTCNAANAGGGTGGGAAAGTCC GAGCCTTANGACCCAGTTTCAGTTCTG GTTTCTTCCCTCCTGANCACCATCGGT TGTTAGTTGCCTTGAGTTGGGAACGTT TGCATCGACACCTGTAAATGTATTCAT TCTTTAATTTATGTAAGGTTTTNTGTNC TCAATTCTTTAAGAAATGACAAATTTT GGTTTTCTACTGTTCAATGAGAACATT AGGCCCCAGCAACACGTCATTGTGTAA ANAAATAAAA
HSC-DD-182	CGATGGCTCCATCCTGGCCTCACTGTC CACCTTCCAGCAGATCGGCTCAGCAAG CAGGAGTAGGATGAGTCTGGCCCCTCC ATCGTGCACCGCAAATGCTTCTAGGCG GACTGTTTTACACCCTTTCTTTGACAA AACC

HSC-DD-089	CNNATGCTACATGCTGNAGGATGCCTA
	AGGCTGCCCCCCACCATCCCCTGGCTC
	TGCTGNCCGGANCAAATTGCTTCCAGA
	TGTGACTTTGGAACCTTCNCACCCCTN
	ACCCNACCNNTCTCNAGAANNTCTTTT
	ATTTAAAGGAGGAAANANNACATCCA
	AGAAAANGGGGGGAGGGGGGATGGA
	AANNCGCATCCCCTTTCTAGCCAGCTG
	TTCCCAAAAGGTACCCTTCCTCTCTGC
	TGCTCCCCAAACNCAAANCCCACTTCN
	GANCCTCCACCTAAANCATCANGCAA
	GTCACNTACACCCTGTTTANCCCCCNA
	CTCTCTGCTTATACCCNGGAACAATTN
	NTGCTCG

-37-

HSC-DD-151	CGATGGTGGGGATCTTACTGGGGAAG
	AGGAAGGACCATTAGCACACCATCAT
	GATGTCAGATGACAAAATGGAAGCCA
	AGACACCTTGAAGGTGACTTTCTAGGA
	AGGTCTTAAGCATGTAATGTCCCTTTA
	TCAGAGGGAAGGGGACAAACTCAGGG
	CAGCCCTGTCCAGGTAGAAATATTTTT
	GCCCCCCTGTCTGATGTTGATGAGGGG
	TCATACCANCCAGGGAGACCCTCTGG
	GAGGAAGCTGCCACACACAANGACTC
	TGGAAGTATCCAGATGTGAGCCCAGC
	CAGGGTCCTATGGTTCCAAATCTGAAN
	AAAAGGTTTTTCACACACTCCTTGCTT
	TCTGCTAAGATAANAAAGGCGTCACTC
	TGCCAGAGTGTGACTTTTTACAGATTA
	AATAAAGCTGTTAT
HSC-DD-013	GATCTACTCCATTCCCCTGGAAATCAT
	GCAGGGCACCGGGGGTGAGCTGTTTG
	ATCACATTGTCTCCTGCATCTCCGACT
	TCCTGGACTACATGGGGATCAAAGGC
	CCCGGATGCCTCTGGGCTTCACCTTCT
	CGTTTCCCTGCAAGCAGACGAGCCTAT
	ATTGCGGAATCTTGATCACGTGGACAA
	AGGGATTCAAAGCCACCGACTGTGTG
	GGTCACNATGTANCCACTTTACTGAG
HSC-DD-029	GATCTGAGTTCGAGGCCAGCCTGGTCT
	ACAGAGTGAGTTCCAGGNCAGCCAGG
	NCTACACAGAGAAACCCTGTCTCGAA
	AAAACAGAAAGAGA

HSC-DD-034	CTTTCATTAAAAAGAAACCAGGGGCT GGANAGATGGCTCAGTGGTTAAGAGC ACCAACTGCTCTTCCCGAAGGTCCTAA GTTCAAATCCCAGCAACCACATGGTGG CTAACAACCACTCGTAATGAGATC
HSC-DD-082B	ATCGCNTGGCTCTCCTGNGGCCTGGCN TACGACNNGAAAAGGAGTGTCCACGG CTGCTGTCGNGGCCACGATTAATTAAA ACTGAAGTACCGAGGNTNCCCCAGNG NCNGANTGTGGGGTCNNGCCNTTCNT GNTCCACAANCCAACTTGGCAGACGC TTACTGTNCTGTCAACTNTCNNNNGAA TACCNCCACCCNCATGCTAAAATGATG ACTGACGTTAANCCATGCTGGT
HSC-DD-084	CGATGACAAAGGAGTCCTGAGGCAGA TTACTCTGAATGACCTTCCTGTCGGAA GATCAGTGGACGAGACACTGCGTTTG GTTCAAGCCTTCCAGTACACTGACAAG CATGGAGAAGTCTGCCCTGCTGGCTGG AAACCTGGTAGTGAAACAATAATCCC AGATCCAGCTGGAAAACTGAAGTATTT CGACAAGCTAAACTGAAAAGTACTTC AGTTATGATGTTTGGACCTTCTCAATA AAGGTCATTGTG

-39-

HSC-DD-128	CGATGCTGAATAAGCTCCTCAAAAAGT GGTAAATTTAACCTTTTNAAAAAACAA GCTTTCTCTGTACAGCTCTGGCTGTTTT GTTCTGGAATACATTCTGTAGAATTGT CTGGCCTCTAACTTGGAGATCCAACTC CCTCTGCCTCTTGAGTGCTGGGATTAA TGGCATGTGACACTGT
HSC-DD-140	CGATGACCTCATGCCGGCCCAGAAGT GAAGCCTGGCCCTCGCCACCATCAGG CTGCCGCTTCCTAACTTATTAACCGGG CAGTGCCCGCCATGCATCCTTGANGTT TGCCGCCTGGCGGCTGAGCCCTTAGCC TCGCTGTAGAGACTTCTGTCGCCCTGG GTAGAGTTTATTTTTTTGATGGNTAAN CTGTTGCTGACACTGAAAATAANCTAG GGTTT
HSC-DD-148	CGATCAATGAAAAGATGACGAGTTTCT TTCAAATGGGCAGTTACTCCCTGATAA CTTCATAGCTGCCTGCACAGAGAAGA AAATCCCTGTTGTGTTTAGACTACAAG AGGGTTATGATCATAGCTACTACTTCA TTGCAACTTTCATCGCTGACCACATCA GACACCATGCTAAGTACCTGAATGCAT GANAAGCCTCAGCCAAGAGAATCTCA TCAGGAGGCCGGAAGGGAATCAACAG GAGTGCTGACTTCCTCGCAGAAGATCA TGCTCCTGCAGCTGAATCGCTTTTCTG AATAAATAT

HSC-DD-176	CGATGTNTACTTCATTGCCACCCTGTC ANTCCTCTGGAAGGTGTCCGTCATCAC CTTGGTCAGCTGTCTCCCCCTCTATGT CCTCAAGTACCTGCGGAGACGGTTCTC CCCACCCAGCTACTCGAAGCTCACTTC CTAAGCTGCAGGGCTGCCTCGGGCAG GGCCTCCGGCCTCCGGCGCTCTCCCAG GAGGAGGTCAAGTTCCACACGCACGA GCCGCCTCTGCTGGACGGTGCAGTCAT GGCTGGCACATGAGGCTTCGCTGAGG CGACACTGGGCACCTAATGGGGATGG AACATTGGTGGAACCGGAGGGAGGGA CCTGAGAGCTGTACCTATCAGAACCTT GGGTGCTAAGCTGTGCTGAGGGGGAA GACGTGGGACCGGATGGCCCGTCTGA GGTTTGTGGGGTCACTGTGCAAGCTTC CTTATGGTTTGAACCTCTTGTCATGTG ATAAAAGT
HSC-DD-178	CGATTTACGTATTTGACTGAAATGAAA GTTCCACTAAACGGTATTTGCTCTTGT GATATGTGGCACATTGTGATATTTTCT TAGTCTGTTCTGTTTCATTTAAAAAAT AAAACTGCTGAT
HSC-DD-180	CCGATGTNCGATAATAGTAAATACCTT AATTANTTAAATAATTCATTGNATTGT TTCAGAGACGTTTGGAAATTACTGTAT ACATTACAACCTAATGACTTTTGTAT TTTATTTTTCAAAANAAAAGCTTA

HSC-DD-186	CNTTNGNNNNTCCNTNCATCNCNGCN GTNTGAGTCCCNCCCAANNAGTCCATC CAANANCCANNGCATNNCAGCTTTAT CATGACAACAAANTGGAGNAAGAAGA AGATGAGTTTCGGCCACTGTTGAGGCA AATCNNTGNNNANTCNTAATANACAC CTGGTCCGCTCATCCTTCAACGTTGTT NTNTANAANTTACCTCCCAGTAGAAA NGCTAGCAANTTTNACCTGCCACNGGT TNTA
HSC-DD-191	CGATCAGATGTCACGCGGGACACANC NCCGCCNCAGTNAATGGNAATATATTT GCATGTTACCCCAAATTANCTTCTNTG CATNGAACATANGTANGTGTCTTTGGG GACACGTGTGTTCTACTAC

-42-

HSC-DD-158	CGATTTACAAATGAACAANCAAGATT ACATATANTGAAAATCCACGCAGGAC CTATTACANAGCATGGTGAAATAGATT ATGAAGCAATTGTAAAGCTTTCAGATG GCTTTAATGGAGCATGACCTGACAAAT GTTTGTACTGAAGCAGGTATGTTTGCA ATTCGTGCCGATCATGATTTTGTANTT CAGGAAGACTTCATGAAAGCAGTCAN GAANGTGGCTGACTCCAAGAAGCTGG AGTCCAAGCTGGACTACAAACCTGTGT GATTCACTANNAGGGTTTGGTGGCTGC ATGACAGACATTGGTTTAATGTANACT TAACNGTTANNGAAACTAATGTANNT ATTGGCAATGANCTTATTANAAGTGAA TANACATGTG
HSC-DD-099	CGATGTTTTTAATTAAGAAGAAATTCA CTTTCTCATTACCTATGAATCTGTGCC AGGGCAGGTGATTTTTGAGTATGAGA ACTTTGTCCTCTCCACAGTTGTCACAA AAATGGTTCCTTCTCATTGAACTATTG TGGCATGCTAATTAAGAAGTGAGTGA CCACTTGGGAGGCAGAGGCAGGTGGA TTTCTGAGTTTGAGGCCAGCCTGGTCT ACAAAGTGAGTTCTAAGACAGCCAGG GCTATACAGAGAAACC

-43-

HSC-DD-222	CCAAGNAATATGGTCTAATCAAAGGT CGTCTGTCTGCTTTTGATTGTCTACATC ACAGCAATCCCTGGGAATTTCTATCCA TTTTAAATGCNGCCGCTTTCATCTGTTT AGCCAGCACACCCAATGGTTTCACTAA CTAGCCCAGTTGACCTTTTGGAAGTTT GAGCCTTGAGCACCTTCAACAAAATTG AGCACTCTGATTAGGATATCCACTTTG CAAATAAAACCAAATGTTTTGTCAAC
HSC-DD-104	CGATGAGGGGAAGATGACCTGGGCCG GGGAGGCCATCCCTTATCCAAGATCAC AGGGAATTCTGGGAAGAGGTTGGCCT GTGGCATCATTGCACGCTCTGCCGGCC TTTTCCAGAACCCCAAGCAGATCTGCT CCTGTGATGGCCTCACTATCTGGGAGG AGCGAGGCCGGCCCATTGCCGGTCAA GGCCGAAAGGACTCAGCCCAACCCCC AGCTCACCTCTAAACAGAGCCTCATGT CAGGTTATTTGGTCCTCGTAGCTGAAC ATCTTCTTGCAGAGGGAGCTGCNGGCC CTTGCTTGTACAGGCCTAAGTACAGGG CAGATAAGTGCTGTAGCCTGAACAAA TTAAATTGTTAC

-44-

HSC-DD-172	CGATTAGCTGNGGTCTCTAGGANATAC TCGTCACTATATGAGCTCAGGANGCCA GCTCTTAGTAGCTCTGAANCAGGTGAA GAATCCTCCTCTGAGGAAACAGACTG GGAGGAAGAAGCAGCCCATTACCAGC CAGCTAATTGGTCAAGAAAAAAGCCA AAAGCNGCTGGCGAAAGTCAGCGTAC TGTTCAACCTCCCGGCAGTCGGTTTCA AGGTCCGCCCTATGCGGAGCCCCCGCC CTGCGTAGTGCGTCAGCAATGCGCAG AGGGGCAATGCGCAGAGAGGTGCGCA GAGGGGCAGTGCGCAGAGAGGTGCGC AGAGAGGCAGTGCGCAGAGAGGCAGT GCGCAGACTCAT
HSC-DD-169	CGATTTCTAAATCAGTCTCGCCTGTGC TAGGATGACCGGTAATGAGCCTGTTTA AAATAAGACTTAAAAGTGTCGTGCGTT GGCCGGGCGGTAGGGGCGCATGCCTT TAATTTCATAACTTGGAGGTAGAGACA GGCGGATCTTTGTGAGTTCAAGGTCAG CCTGGTGTACAGAGTGACTTCCAGAAC AGCCAGGGCTGTTAAACAGAGAAAC
HSC-DD-003A	TTGTTTTGTTNTTCAGATAGGGTCTTAC ATATCCCATGCTGGTCTCAAACTCACA TTATGCATGCGGGGAAAGCCATTTACT GACTGATATACCCCTGGCCCTAAGATA GATC

-45-

HSC-DD-092	CGATCGTCGTTCTGGTAAGAAGCTGGA AGATGGCCCCAAGTTCCTGAAGTCTGG CCATTTAAGTTTAATAGTAAAAGACTG GTTAATGATAACAATGCATCGTAAAAC
HSC-DD-114	CGATCGTCGTTCTGAGTAANAAGCTGG AANANGGCCCCAAGTTCCTGNNGTCT GGCGATGCTGCCATTTAAGTTNANNAG ANANAAGACTGGCTNATGATAACAAT GCANCNTAAAACCTTCAGGNAGGNAA CGAATGTTGTGGACCATTTTTTNTGNG TGTGGCAGTTTNAAGTTATNAAGNTTT CAAAANCANTACTTNTTAANGGGAAC AACTTGACCCATCANCTGTCACAGAAT NTTGANGACCATTAACAC
HSC-DD-213A1	NCTACGATCATCTAGATCTACTAGACC TACNACNAGACCATGGGCCAAANATG GTCGACCTGCAAACTTGCAAGGTTTAT TTTANATACACATTATGGCGTTTTATN TTTTGTAATTCTAAGTTGTAATTCAGCT TTTAACAAATCTTTTT
HSC-DD-213A1'	CCAAGNANATCNAGACTACTAGACCT ACTACNAGACCATNGGNCAAACATGG TCGACCNNCAAACGNATANGTATATTT NANATACACANANATAGCGTTNTATG TCTNGTAATTCTAAGTNGTANATCANC TATTANCAAAATCTTTNTTT

HSC-DD-155	CGATGGAAGTTCTGCTGAGCCCTTCTG ACGTAACCCTGGCNATGGCTAACACTG TCCTTCCTGCAATGTTCNTGGTGGACA CANCTTCTCTGGANATACCCTGAANGT GGCACGCCCTGTTCCAGCCCACCTGGT GTGCACTTTTGCCCTCTTTACCTCATT ANTAAATGTTTTCNTGCTCCTAATG
HSC-DD-212	CTNAGNAAGGANCTGTACTTCGTATTG CAAGGCAGTCTCTTGTGTCTTCTTAGA GTGTCTTCCCCATGCACAGCCTCAGTT TGGAGCACTAGTTTATAATGTTTATTA CAATTTTTAATAAATTGANTAGGTAGT A
HSC-DD-090	TCNTCNTTCTGGTAAGAACTGGAATAT GGCCCCAAGTTCCTGAAGTCTGGCGAT GCTGCCATTGTTGATATGGTCCCTGGC AANCCCATGTGTGTTGAGAGCTTCTCT GACTACCCTCCACTTGGTCGCTTTGCT GTTCGTGACATGAGGCAGACAGTTGCT GTGGGTGTCATCAAAGCTGTGGACAA AAANGCTGCTGGAGCTGGCNAAGTCA CCAAGTCTGCCCANAAAGCTCAGAAG GCTAAATGAATATTACCCCTAACANCT GCCACCNCANTCTTAATCAGTGGTGGA AGAACGGTCTCAGAACTGTTNGTCTCA ANTGGCCATTTAAGTTTAATANTAAAA GACTGGTTAATGATAAC

-47-

HSC-DD-173	CGATCNTCGTTCTGGTAAGANNCNGG AACATGGCCCCAAGTTCCNGANNTCTG GCGANGCNGCCANTGTTGATATGGTCC CTGGCAAGCCCATGTGTNTTGAGAGCT TCACNNACNACCCTCCANTTGGTCGCT TTGCTGTTCGTGACATGAGGCAGACAG TTGCTGTGGGTGTCANCAAANCTGTGG ACAANANGGCTGCTGGAGCTGGCAAG NTCACCAANTCTGCCCAGAAAGCTCA GAATGCTAAATNAATATTACCCCTAAN ACCTGCCACCCCAGTCNTAATCAGTGG TGGAATAACNGTCTCAGAACTGTTTGT CNCAATTGGCCANTTANGTTTAATNAT ACAAGACTG
HSC-DD-249	GNNNNNNNNNNNCNANGAAAAAGAG GTGAAAAATGCTTGGCTCTAGCTGATG ACAGAAAGCTGAAATCCATCGCCTTCC CATCCATTGGCAGCGGCAGGAACGGG TTCCCGGAAGCAGACAGCGGCCCAGC TCATTCTGAAGTGCCATCTCCAGCTAC NTTGTCTCCACGATGTCCTCCTCCATC AAAACTGTGTACTTCATGCTTTTTGAC AGTGAGAGCATAGGTATCTATGTGCA GGAAATGGCCAAGCTGGACGCCAACT AGGCCAGTGATCCCTAGAGCCAGCAC ATGCGGTGTCCCCCA

HSC-DD-250	CTNANGAAAGCTGCTGGGGCNCCCTG ACATCACTCATCACTCACTATGCTACC AATTCTATTTATTTCGGAATTACAAGA TATCGGGAATCTCTCTGCAGGCTGGAC TGGCAGGCTGTGGGGTGGGCGGGACA CGGCTCTTAACATTTNCAGAGGGAAAC GCGCANATGTCCAAAAGTCTAAATAA ATGCATTCAGAGGTTTNTGGGGTCCAT GGCCAAGTGGAGTTCCCCCNCAGGGG GAGGTGGGGTAAGTGCCTCCAGGAAG GCAGGCAGCCTGCCTTANACTTGCANC CCGGNTGTGGGAATGAATCATTGGAG TAATAAACT
HSC-DD-108	CGATGCCAATGGCATCCTCAATGTTTC TGCTGTAGATAAGAGCACAGGAAAGG AGAAAGTCTGCAACCCTATCATTACCA AGCTGTACCAGAGTGCAGGTGGCATG CCTGGGGGAATGCCTGGTGGCTTCCCA GGTGGAGGAGCTCCCCCATCTGGTGGT GCTTCTTCAGGCCCCACCATTGAAGAG GTGGATTAAGTCAGTCCAAGAAGAAG GTGTAGCTTTGTTCCACAGGGACCCAA AACAAGTAACATGGAATAATAAAACT ATTTA

-49-

HSC-DD-116	CGATGAAGATGAGGTCACTGCAGAGG
	AGCCCAGTGCTGCTGTTCCTGATGAGA
	TCCCCCCTCTGGAAGGCGATGAGGATG
	CCTCGCGCATGGAAGAGGTGGATTAA
	AGCCTCCTGGAAGAAGCCCTGCCCTCT
	GTATAGTATCCCCGTGGCTCCCCCAGC
	AGCCCTGACCCACCTGGATCTCTGCTC
	ATGTCTACAAGAATCTTCTATCCTGTC
	CTGTGCCTTAAGGCAGGAAGATCCCCT
	CCCACAGAATAGCAGGGTTGGGTGTT
	ATGTATTGTGGTTTTTTTGTTTGTTTTA
	TTTTGTTCTAAAATT

-50-

HSC-DD-166	CGATGCCAATGGCATCCTCAATGTTTC
	TGCTGTAGATAAGAGCACAGGAAAGG
	AGAACAAGATCACCATCACCAATGAC
	AAGGGCCGCTTGAGTAAGGAAGATAT
	TGAGCGCATGGTCCAAGAAGCTGAGA
	AGTACAAGGCTGAGGATGAGAAGCAG
	AGAGATAAGGTTTCCTCCAAGAACTCA
	CTGGAGTCCTATGCCTTCAACATGAAA
	GCAACTGTGGAAGATGAGAAACTTCA
	AGGCAAGATCAATGATGAGGACAAAC
	AGAAGATTCTTGACAAGTGCAATGAA
	ATCATCAGCTGGCTGGATAAGAACCA
	GACTGCAGAGAAGGAAGAATTTGAGC
	ATCAGCAGAAAGAACTGGAGAAAGTC
	TGCAACCCTATCATTACCAAGCTGTAC
	CAGAGTGCAGGTGGCATGCCTGGGGG
	AATGCCTGGTGGCTTCCCAGGTGGAGG
	AGCTCCCCCATCTGGTGGTGCTTCTTC
	AGGCCCCACCATTGAANAGGTGGNTT
	AAGTNATCCANNAAGAAAGGNTNCCT
	TTTTTTCCAAAGGGANCCAAAAAAGTA
	ANATGGATAATAAAACCTATTTAATT

-51-

HSC-DD-184	CGATGCCAATAGNANCCCAANTNTCT GCNGTNGATAAGACACANGAAAAGAG AACAAGATCACCATCACCAATGACAA GGGCCGCTTGAGTAAGGAAGATATTG AGCGCATGGTCCAAGATCAATGATGA GGACAAACAGAAGATTCTTGACAAGT GCAATGAAATCATCAGCTGGCTGGAT AAGA
HSC-DD-101	CGATTAGCGGAGGTCTCTAGGAGATA CTCGTCACTAGATGAGCTCAGGAAGCC AGCTCTTAGTAGCTCTGAAGCAAGTGA AGAATCCTCCTCTGAGGAAACAGACT GGGAGGAAGAAGCAGCCCATTACCAG CCAGCTAATTGGTCAAGAAAAAAGCC AAAAGCGGCTGGCGAAAGTCAGCGTA CTGTTCAACCTCCCGGCAGTCGGTTTC AAGGTCCGCCCTATGCGGAGCCCCCG CCCTGCGTAGTGCGTCAGCAATGCGCA GAGGGGCAATGCGCAGAGAGGCAGTG CGCAGAGAGGCAGTGCGCAGACTCAT TCATT
HSC-DD-017	TCTCTGTATAACCCTGGATGTCCTGGA ACTCACTTTGTAGACCAGGTTGGCCTC GAACTCAGAAATCCGCCTGCCTCTGCC AAGCGCTGGGATTAAAGGTGTGCGCC ACCACACCCGGCAGGTAATTTTTTTCT TTTTAAAGATTTATTATGTATACAGGT TCTGCCTACATGTGTACCTGCCGGCCA GAAGAGGGCATCANATC

| HSC-DD-026 | GATCTTTGTAGGCACAAAATGAATCCC GCACCTGGTGACCCATGATGCTCGTAC TATTCGGTACCCTGATCCCCTCATCAA GGTGAACGACACCATTCAGATTGATTT GGAGACAGGCAAAATAACTGACTTCA TCAAGTTTGACACTGGGAACCTGTGTA TGGTGACTGGAGGTGCTAACTTGGGA AGAATTGGTGTAATCACCAACAGAGA GAGACATCCCGGCTCTTTTGATGTGGT TCATGTGAAAGATGCCAATGGCAACA GCTTTGCCACTCGGCTGTCCAACATTT TTGTTATTGGCAAGGGTAACAAACCAT GGATCTCTCTTCCCAGAGGAAAAGGA ATCCGCCTCACCATTGCTGAAGAGAGA GACAAGAGGCTTGCGGCCAAACAGAG CAGTGGGTTGAAATGGTCTCCTAGGAG ACATGCCTGGAAAGTTGTTTTGTACAA CCTTTCTCAGGCAACATACATTGCTAG AATTAAACAGCCATG |
| HSC-DD-064 | CGATCGAGAGGGCAAACCACGGAAGG TGGTTGGTTGCAGTTGCGTAGTGGTTA AGGACTATGGCAAAGAATCTCAGGCC AAGGATGTCATCGAGGAAATACTTCA AGTGCAAGAAATAAATAAATTTTGGCT GATT |

-53-

HSC-DD-066	ATTCCAGATGAGGACCACAAGCGACT CATTGATTTACATAGTCCTTCTGAGAT TGTTAAGCAGATTACTTCCATCAGTAT TGAGCCGGGAGTTGAGGTTGAAGTCA CCATTGCAGATGCCTAAGACAACTGA ATAAATCG
HSC-DD-041	GATCTATACAGTCGGGAAACGCTTCAA GGAAGCAAATAACTTCCTGTGGCCCTT CAAGTTATCTTCCCCACGAGGTGGGAT GAAGAAAAAGACAACTCACTTTGTAG AAGGTGGAGATGCTGGCAACAGGGAA GACCAGATAAACAGGCTTATTAGACG GATGAACTAAGGTGTCACCCATTGTAT TTTTGTAATCTGGTCAGTTAATAAACA GTC
HSC-DD-111	CGATGTGGCCAAAGTCAATACCCTGAT AAGGCCCGACGGAGAGAAGAAGGCGT ATGTTCGCTTGGCTCCTGATTATGATG CCCTAGATGTTGCCAACAAGATTGGGA TCATCTAAACTGAGTCCAGATGGCTAA TTCTAAATATATACTTT
HSC-DD-028B	GATCTGGAACCATAGATGCGAGCATC AGCAACAGAATACAAGAAATGGAAGN GNGAATCTCAGGTGCAGAAGNTTCCA TAGAGAACATCG

-54-

HSC-DD-142	GCGATGCAAAATCCTTAATANAATTCT TGCTAACCGAATCCAAGAACACATTA AAGCAATCATCCATCCTGACCAAGTAG GTTTTATTCCAGGGATGCNGNGATGGT TTAATATATGAAAATCCATCAATGTAA TCCATTNTATAAACAANCTCAANGACA NAAACCACATGATCATCTCGTTAGNTG CAGAAAAGCATTTGACAAGATCCAA CACACATTCGTGATAANAGTTTTGGNA AGATCAGGAATTCAAG
HSC-DD-095	CGATNNACCCGCTCTACCTCACCATCT CTTGCTAATTCAGCCTATATACCGCCA TCTTCAGCAAACCCTAAATNAGGTATT AAAGTAAGCATCNAGAATCANCCATA CTCAACGTNACGTCAAGGTGTACCCAA TGNAATGGGAAGAAATGGGCTACATT TTCTTATANAAGAACATTNCTATACCC TTTNTGAAACTAA

Table 3 presents the expression patterns of the differentially expressed bands set

5   forth in Table 2. The band fragment length (size) in Table 3 is the length before

unwanted terminal sequences were removed. Table 3 also presents the results of a

GenBank Search and analysis of the sequences of Table 2.

Summary of Known Genes from Mouse HSC Differential Display (I)

Items No.	Size (bp)	Enzyme	N1N2 (oligo-dT)	Poly(A) Sign	Lin-	LRH	LRH4N	LRBRH	Gene Bank Search & Analysis
HSC-DD-006	213	Bgl II	AC	fair	0	3+	/	+	mouse homeobox protein
HSC-DD-285	158	Xba I	GG	good	±	+	+	±	human homeobox gene regulator
HSC-DD-007B	213	Bgl II	AC	fair	±	2+	/	±	human zinc finger protein 10
HSC-DD-238	363	Xba I	AG	good	3+	0	3+	3+	mouse cell division control protein 19
HSC-DD-206	123	Xba I	AC	good	3+	0	2+	+	human HS1 heamalopoietic protein
HSC-DD-214	192	Xba I	AC	fair	+	2+	0	3+	mouse plm-1 proto-oncogene
HSC-DD-035	151	Bgl II	AC	fair	±	2+	/	+	mouse thyroid hormone receptor
HSC-DD-129	234	Cla I	AC	poor	0	3+	0	0	mouse Inositol 1,4,5-trisphosphate receptor
HSC-DD-040	220	Bgl II	AC	fair	+	2+	0	0	mouse G protein beta-36 subunit
HSC-DD-011	173	Bgl II	AC	good	±	±	/	2+	mouse ras-related YPT1 protein
HSC-DD-121	186	Cla I	CT	poor	0	3+	±	±	human TBP-associated factor 170
HSC-DD-015B	133	Bgl II	AG	poor	0	3+	/	+	mouse HMG1-related DNA binding protein
HSC-DD-039	206	Bgl II	AC	fair	2+	4+	/	4+	mouse TAX responsive element binding protein 107
HSC-DD-042	235	Bgl II	AC	fair	±	0	/	+	mouse retinoblastoma binding protein isoform III
HSC-DD-256	272	Xba I	AA	poor	0	2+	±	0	Rat androgen-binding protein
HSC-DD-045	270	Bgl II	AC	good	±	2+	/	±	similar to Rat cca2
HSC-DD-068	164	Cla I	AC	fair	+	4+	4+	4+	mouse jerky mRNA
HSC-DD-143	350	Cla I	AG	fair	±	2+	±	±	similar to human memd
HSC-DD-263	292	Xba I	AT	good	0	2+	±	0	mouse interleukin 5
HSC-DD-239	156	Xba I	CA	good	±	3+	3+	+	human CD9
HSC-DD-261	115	Xba I	AA	good	0	+	0	0	mouse germline kJM
HSC-DD-028A	95	Bgl II	AC	good	+	4+	/	.	mouse chaperonin containing TCP-1 e subunit
HSC-DD-021	143	Bgl II	AG	lar	±	+	/	2?	mouse calreticulin
HSC-DD-025	326	Bgl II	AG	good	±	2?	/	2?	mouse metallothionein I

Summary of Known Genes from Mouse HSC Differential Display (II)

Items No.	Size (bp)	Enzyme	N1N2 (oligo-dT)	Poly(A) Sign	Expression pattern				Gene Bank Search & Analysis
					Lin'	LRH	LRH4X	LRBRH	
HSC-DD-077	203	Cla I	AC	good	+	2+	2+	3+	Rat matrin cyclophilin
HSC-DD-200	450	Cla I	AA	fair	+	±	2+	+	mouse G-utrophin
HSC-DD-245	272	Xba I	CA	fair	3+	±	3+	2+	rat basement membrane-associated chondroitin
HSC-DD-226	387	Xba I	AC	good	±	3+	±	0	mouse cytoplasmic g-actin
HSC-DD-182	149	Cla I	GC	poor	±	3+	±	+	mouse A-X actin
HSC-DD-089	364	Cla I	AC	poor	+	3+	2+	+	mouse TIE receptor tyrosine kinase
HSC-DD-151	424	Cla I	GA	good	0	+	2+	±	rat elk, brain-specific receptor tyrosine kinase
HSC-DD-013	248	Bgl II	AC	fair	±	2+	/	3+	mouse hexokinase
HSC-DD-029	103	Bgl II	AC	fair	0	+	/	0	mouse bruton agammaglobulinemia tyrosine kinase
HSC-DD-034	140	Bgl II	AC	fair	0	2+	/	2+	mouse spermine synthase
HSC-DD-082B	244	Cla I	AC	fair	+	4+	2+	2+	mouse stearoyl-CoA desaturase (SCD2)
HSC-DD-084	261	Cla I	AC	good	±	+	±	2+	mouse antioxidant enzyme AOE 372
HSC-DD-128	189	Cla I	AC	fair	0	3+	3+	±	mouse casein kinase II beta chain
HSC-DD-140	229	Cla I	AG	good	±	0	0	+	mouse creatine kinase B
HSC-DD-148	313	Cla I	GA	good	+	+	2+	±	human esterase D
HSC-DD-176	470	Cla I	CG	fair	±	3+	+	0	mouse putative E1-E2 ATPase
HSC-DD-178	130	Cla I	GC	good	±	3+	0	+	mouse aspartate aminotransferase
HSC-DD-180	142	Cla I	GC	good	+	+	0	+	mouse tyrosylprotein sulfotransferase-1
HSC-DD-186	252	Cla I	GC	poor	±	+	2+	2+	mouse ubiquitin-conjugating enzyme E214K
HSC-DD-191	136	Cla I	AA	fair	0	±	3+	2+	mouse b-1,4-galactosyltransferase
HSC-DD-158	391	Cla I	GT	fair	+	3+	0	+	spermophilus tridecemlineatus 26s proteasome
HSC-DD-099	265	Cla I	CC	fair	±	3+	0	±	mouse proteasome epsilon chain precursor
HSC-DD-222	270	Xba I	AC	good	0	2+	3+	+	Rat 3-hydroxyso- butyrate
HSC-DD-104	368	Cla I	CC	fair	0	±	+	±	human copper chaperone for superoxide dismutase
HSC-DD-172	365	Cla I	CG	fair	±	3+	2+	0	mouse Ercc 4 DNA repair gene
HSC-DD-169	223	Cla I	CG	fair	±	±	2+	0	Cricetulus griseus nucleotide excision repair protein
HSC-DD-003A	148	Bgl II	AC	poor	0	+	/	±	human G rich sequence factor

Summary of Known Genes from Mouse HSC Differential Display (III)

Items No.	Size (bp)	Enzyme	N1N2 (oligo-dT)	Poly(A) Sign	Expression pattern				Gene Bank Search & Analysis
					Lin+	LRH	LRH48	LRBRH	
HSC-DD-092	118	Cla I	CC	fair	+	3+	+	+	mouse elongation factor 1-a
HSC-DD-288	480	Xba I	GC	fair	±	+	+	±	human elongation factor-1-delta
HSC-DD-114	267	Cla I	CA	poor	±	+	±	+	Rat elongation factor-1-alpha
HSC-DD-213	178	Xba I	AC	fair	±	3+	+	+	human splicing factor (SFRS7)
HSC-DD-155	198	Cla I	GT	fair	0	2+	+	0	mouse transcription elongation factor S-II-T1
HSC-DD-212	162	Xba I	AC	poor	0	3+	±	0	mouse translation initiation factor 4E
HSC-DD-090	375	Cla I	AC	fair	±	3+	3+	+	mouse protein synthesis elongation factor
HSC-DD-173	367	Cla I	CG	fair	±	3+	+	0	mouse protein synthesis elongation factor Tu
HSC-DD-249	304	Xba I	CA	poor	4+	+	4+	4+	rat histone macroH2A1.2
HSC-DD-250	356	Xba I	CA	good	+	2+	3+	2+	mouse MER9 processed pseudogene
HSC-DD-108	281	Cla I	GG	good	+	2+	+	2+	mouse heat shock protein 70
HSC-DD-116	326	Cla I	CA	fair	±	2+	0	2+	mouse 84 kD heat shock protein
HSC-DD-166	587	Cla I	AT	good	±	2+	3+	+	mouse heat shock protein 70 cognate
HSC-DD-184	196	Cla I	GC	fair	±	2+	0	±	mouse breast heat shock protein 73
HSC-DD-101	331	Cla I	CC	fair	+	3+	0	±	mouse MHC locus II region
HSC-DD-017	215	Bgl II	AG	good	0	4+	/	0	mouse MHC class III region
HSC-DD-026	505	Bgl II	AG	fair	2+	4+	/	4+	mouse ribosomal protein S4
HSC-DD-064	146	Cla I	AC	good	2+	2+	2+	3+	mouse ribosomal protein S12
HSC-DD-066	150	Cla I	AC	good	2+	3+	2+	2+	mouse ribosomal protein S20
HSC-DD-041	226	Bgl II	AC	good	*	3+	/	3+	mouse ribosomal protein L7
HSC DD 111	161	Cla I	CA	lar	*	*	*	*	rat ribosomal protein L23a
HSC DD 0298	100	Bgl II	AC	lar	*	4+	/	*	mouse LINE-1/L1 element
HSC DD 142	267	Cla I	AG	lar	±	2+	±	±	mouse L1Md A13 repetitive sequence
HSC DD 095	210	Cla I	CC	lar	±	2+	±	/	mouse mitochondrial 12S ribosomal RNA

-58-

As is apparent to one of ordinary skill in the art, this same procedure can be used to identify stem cells genes whose expression levels are associated with stem cell proliferation, dedicated differentiation and survival.

5   Example 2
Method to identify a therapeutic agent that modulates the expression of at least one stem cell gene associated with the differentiation process of a stem cell population.

The methods set forth in Example 1 offer a powerful approach for identifying therapeutic agents that modulate the expression of at least one stem cell gene associated
10   with the differentiation process of a stem cell population. For instance, gene expression profiles of undifferentiated stem cells and partially differentiated or terminally differentiated stem cells are prepared as set forth in Example 1. A profile is also prepared from an undifferentiated stem cell sample that has been exposed to the agent to be tested. By examining for differences in the intensity of individual bands between the three
15   profiles, agents which up or down regulate genes associated with the differentiation process of a stem cell population are identified.

Example 3
Method to identify a therapeutic agent that modulates the expression of at least one stem
20   *cell gene associated with the proliferation of a stem cell population.*

The methods set forth in Example 1 offer a powerful approach for identifying therapeutic agents that modulate the expression of at least one stem cell gene associated with the proliferation of a stem cell population. For instance, gene expression profiles of undifferentiated stem cells and actively proliferating stem cells are prepared as set forth
25   in Example 1. A profile is also prepared from an undifferentiated stem cell sample that has been exposed to the agent to be tested. By examining for differences in the intensity of individual bands between the three profiles, agents which up or down regulate genes associated with the proliferation of a stem cell population are identified.

As is apparent to one of ordinary skill in the art, this same procedure can be used to identify stem cells genes whose expression levels are associated with stem cell dedicated differentiation and survival.

Example 4

5   *Production of solid support compositions comprising groupings of nucleic acids or nucleic acid fragments that correspond to genes whose expression levels are associated with the differentiation, proliferation, dedicated differentiation or survival of stem cells.*

As set forth in Example 1, expression profiles prepared from stem cells at different stages of differentiation, from proliferating stem cells, from stem cells that are

10   dedicated to a differentiation pathway and from stem cells resistant to apoptosis (which may be linked to increased survival) provide a means to identify genes whose expression levels are associated with stem cell differentiation, proliferation, dedicated differentiation and survival, respectively.

Solid supports can be prepared that comprise immobilized representative

15   groupings of nucleic acids or nucleic acid fragments corresponding to the genes from stem cells whose expression levels are modulated during stem cell differentiation, proliferation, dedicated differentiation and survival. For instance, representative nucleic acids can be immobilized to any solid support to which nucleic acids can be immobilized, such as positively charged nitrocellulose or nylon membranes (see Sambrook *et al.*

20   (1989) *Molecular Cloning: a Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory) as well as porous glass wafers such as those disclosed by Beattie (WO 95/11755). Nucleic acids are immobilized to the solid support by well established techniques, including charge interactions as well as attachment of derivatized nucleic acids to silicon dioxide surfaces such as glass which bears a terminal epoxide moiety. At

25   least one species of nucleic acid molecule, or fragment of a nucleic acid molecule corresponding to the genes from stem cells whose expression levels are modulated during stem cell differentiation, proliferation, dedicated differentiation and survival may be immobilized to the solid support. A solid support comprising a representative grouping of nucleic acids can then be used in standard hybridization assays to detect the presence

-60-

or quantity of one or more specific nucleic acid species in a sample (such as a total

cellular mRNA sample or cDNA prepared from said mRNA) which hybridize to the

nucleic acids attached to the solid support. Any hybridization methods, reactions,

conditions and/or detection means can be used , such as those disclosed by Sambrook *et*

5   *al.* (1989) *Molecular Cloning: a Laboratory Manual*, 2nd Ed., Cold Spring Harbor

Laboratory, Ausbel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene

Publishing and Wiley-Interscience. N.Y. or Beattie in WO 95/11755.

One of ordinary skill in the art may determine the optimal number of genes that

must be represented by nucleic acid fragments immobilized on the solid support to

10   effectively differentiate between samples that are at the various stages of stem cell

differentiation, including terminal differentiation, proliferating stem cells, stem cells

dedicated to a given differentiation pathway and/or stem cells with increased survival

rates. Preferably, at least about 5, 10, 20, 50, 100 , 150, 200, 300, 500, 1000 or more

preferably, substantially all of the detectable mRNA species in a cell sample or

15   population will be present in the gene expression profile or array affixed to a solid

support. More preferably, such profiles or arrays will contain a sufficient representative

number of mRNA species whose expression levels are modulated under the relevant

differentiation process, disease, screening, treatment or other experimental conditions. In

most instances, a sufficient representative number of such mRNA species will be about 1,

20   2, 5, 10, 15, 20, 25, 30, 40, 50, 50-75 or 100 in number and will be represented by the

nucleic acid molecules or fragments of nucleic acid molecules immobilized on the solid

support. For example, nucleic acids encoding all or a fragment of one or more of the

known genes or previously reported ESTs that are identified in Tables 2 and 3 may be so

immobilized. Additionally, the skilled artisan may select nucleic acids encoding the

25   protein cell surface markers discussed above at page 8 (*i.e.*, CD 34) in order to help

identify the particular stage of differentiation of a given stem cell population and to

identify agents that are involved in promoting such differentiation. The skilled artisan

will be able to optimize the number and particular nucleic acids for a given purpose, i.e.,

screening for modulating agents, identifying activated stem cells, etc.

-61-

In general, nucleic acid fragments comprising at least one of the sequences or part of one of the sequences of Table 2 can be used as probes to screen nucleic acid samples from cell populations in hybridization assays. Alternatively, nucleic acid fragments derived from the identified genes in Table 3 which correspond to the sequences of Table

5   2 may be employed as probes. To ensure specificity of a hybridization assay using probe derived from the sequences presented in Table 2 or the genes of Table 3, it is preferable to design probes which hybridize only with target nucleic acid under conditions of high stringency. Only highly complementary nucleic acid hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the

10   amount of complementarity which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the difference in stability between the probe:target hybrid and potential probe:non-target hybrids.

Probes may be designed from the sequences of Table 2 or the genes of Table 3 through methods known in the art. For instance, the G+C content of the probe and the

15   probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available in Sambrook et al. (*Molecular Cloning: A Laboratory Approach,* Cold Spring Harbor Press, NY, 1989) or Ausubel et al. (*Current Protocols in Molecular Biology,* Greene Publishing Co., NY, 1995). Any available format may be used in designing hybridization assays, including immobilizing the probes to a solid

20   support or immobilizing the cellular test sample nucleic acids to a solid support.

It should be understood that the foregoing discussion and examples merely present a detailed description of certain preferred embodiments. It therefore should be apparent to those of ordinary skill in the art that various modifications and equivalents can be made without departing from the spirit and scope of the invention. All documents,

25   patents and references, including provisional patent application 60/056,861, referred to throughout this application are herein incorporated by reference.

-62-

What is Claimed Is:

1.   A method to identify an agent that modulates the expression of at least one stem cell gene associated with the differentiation process of a stem cell population, comprising the steps of:

5   preparing a first gene expression profile of an undifferentiated stem cell population;

              preparing a second gene expression profile of a stem cell population at a defined stage of differentiation;

              treating said undifferentiated stem cell population with the agent;

10   preparing a third gene expression profile of the treated undifferentiated stem cell population;

              comparing the first, second and third gene expression profiles; and

              identifying an agent that modulates the expression of a least one gene in undifferentiated stem cells that is associated with stem cell differentiation.

15   2.   A method to identify an agent that modulates the expression of at least one stem cell gene associated with the proliferation of a stem cell population, comprising the steps of:

              preparing a first gene expression profile of a non-proliferating stem cell population;

20   preparing a second gene expression profile of a proliferating stem cell population;

              treating the non-proliferating stem cell population with the agent;

              preparing a third gene expression profile of the treated stem cell population;

25   comparing the first, second and third gene expression profiles; and

              identifying an agent that modulates the expression of a least one gene that is associated with stem cell proliferation.

-63-

3.   A composition comprising a grouping of nucleic acid molecules that correspond to at least part of the sequences of Table 2 or genes of Table 3 affixed to a solid support.
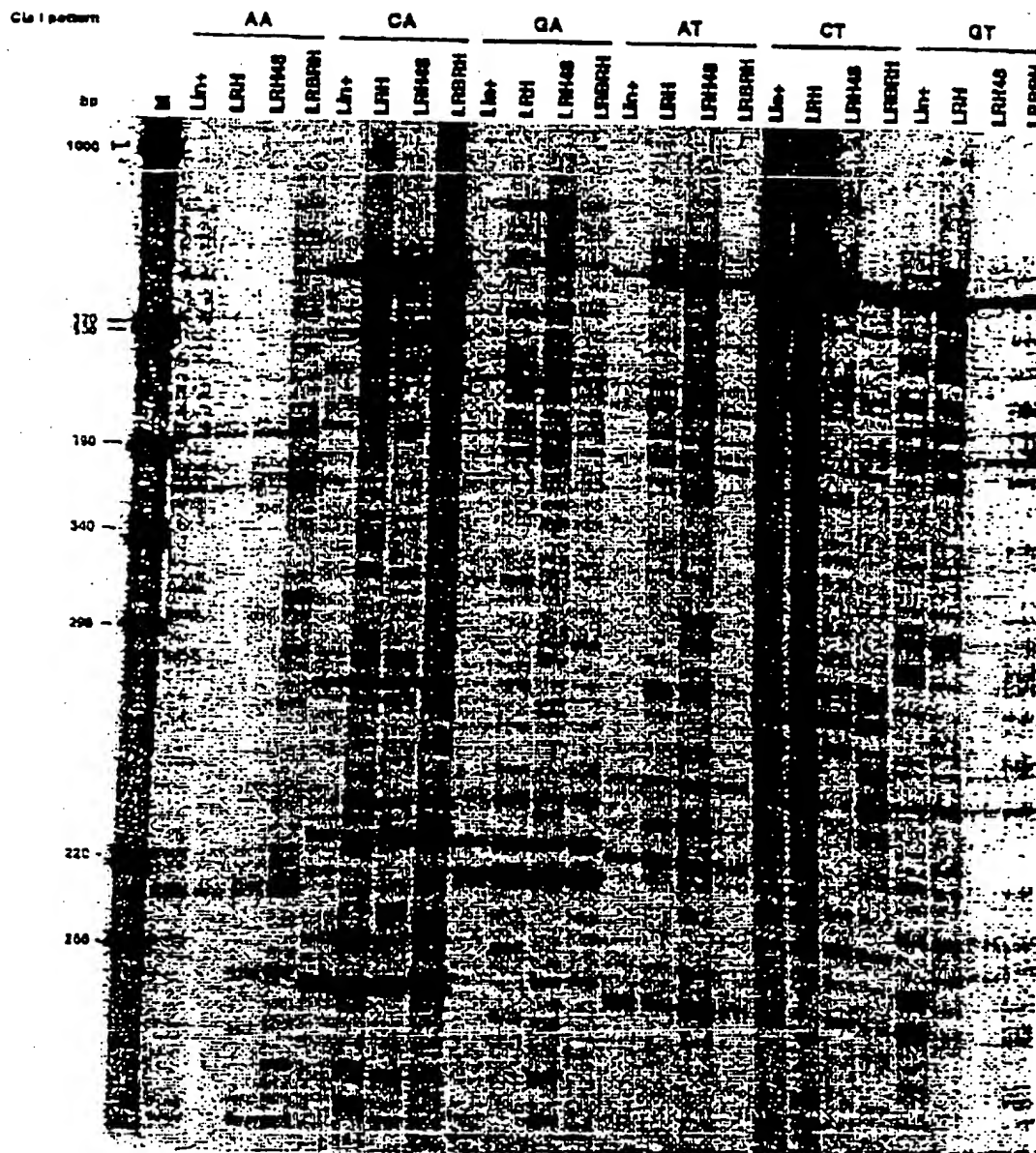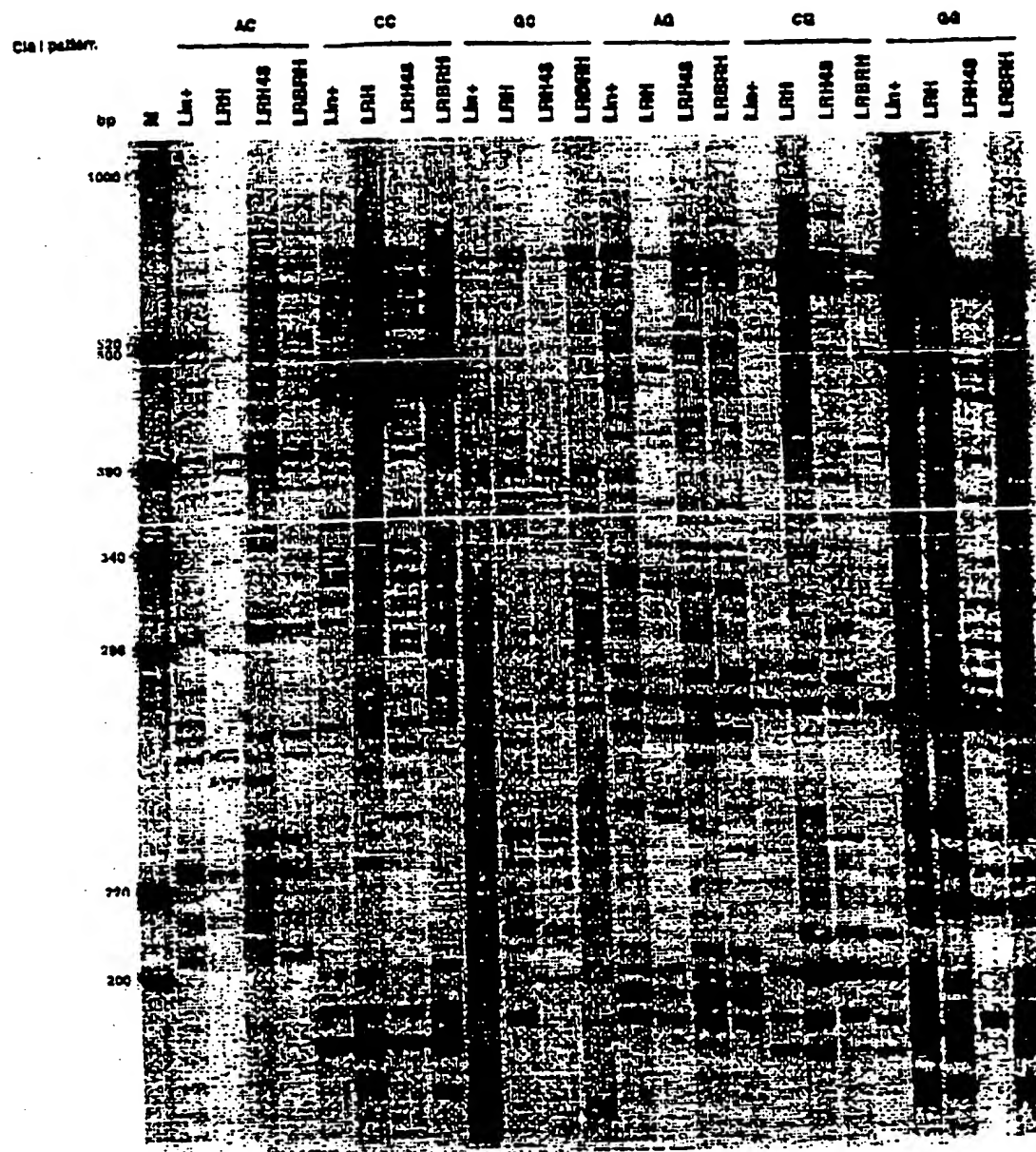
FIG. 1

FIG. 1 (Cont.)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/17283

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C12Q 1/68; C12N 15/12
US CL : 435/6; 536/23.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, Medline, WPIDS
search terms: hematopoietic stem cell, differential display

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TAGOH et al. Molecular Cloning and Characterization of a Novel Stromal Cell-Derived cDNA Encoding a Protein That Facilitates Gene Activation of Recombination Activating Gene (RAG)-1 in Human Lymphoid Progenitors. Biochem. Biophys Res. Commun. 1996, Vol. 221, pages 744-749, especially page 744.	1, 2
X	MOREB et al. Human A1, a Bcl-2-related gene, is induced in leukemic cells by cytokines as well as differentiating factors. Leukemia. July 1997, Vol. 11, Number 7, pages 998-1004, especially page 998.	1, 2

☐ Further documents are listed in the continuation of Box C.	☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 NOVEMBER 1998	24 DEC 1998

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JOHN S. BRUSCA
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/17283

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 3
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

   No sequence listing or computer readable form of sequence listing has been supplied, and claim 3 is drawn to specific sequences that therefore cannot be searched.

3. ☐ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest   ☐ The additional search fees were accompanied by the applicant's protest.
                             ☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)*

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) Internati nal Patent Classification [6] : C12Q 1/68, C12N 15/12	A1	(11) International Publication Number: **WO 99/10535**
		(43) International Publication Date: 4 March 1999 (04.03.99)

(21) International Application Number: PCT/US98/17283

(22) International Filing Date: 21 August 1998 (21.08.98)

(30) Priority Data:
60/056,861   22 August 1997 (22.08.97)   US

(71) Applicant (for all designated States except US): YALE UNI-VERSITY [US/US]; 451 College Street, New Haven, CT 06520 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): LIU, Meng [CN/US]; Apartment 7C, 564 Prospect Street, New Haven, CT 06511 (US). BASKARAN, Namadev [IN/US]; 750 Whitney Avenue, New Haven, CT 06511 (US). WEISSMAN, Sherman, M. [US/US]; 459 Saint Ronan Street, New Haven, CT 06511 (US).

(74) Agent: ADLER, Reid, G.; Morgan, Lewis & Bockius LLP, 1800 M Street, N.W., Washington, DC 20036 (US).

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: A PROCESS TO STUDY CHANGES IN GENE EXPRESSION IN STEM CELLS

(57) Abstract

The present invention includes a method to identify stem cell genes that are differentially expressed in stem cells at various stages of differentiation when compared to undifferentiated stem cells by preparing a gene expression profile of a stem cell population and comparing the profile to a profile prepared from stem cells at different stages of differentiation, thereby identifying cDNA species, and therefore genes, which are expressed. The present invention also includes methods to identify a therapeutic agent that modulates the expression of at least one stem cell gene associated with the differentiation, proliferation and/or survival of stem cells.

-1-

A PROCESS TO STUDY CHANGES
IN GENE EXPRESSION IN STEM CELLS

<u>Technical Field</u>

This invention relates to compositions and methods useful to identify agents that modulate the expression of at least one gene associated with the differentiation, proliferation, dedication and/or survival of stem cells.

5   <u>Background of the Invention</u>

The identification of genes associated with development and differentiation of cells is an important step for advancing our understanding of hematopoiesis, the differentiation of hematopoietic stem cells into erythrocytes, monocytes, platelets and polymorphonuclear white blood cells or granulocytes. The identification of genes
10 associated with hematopoiesis is also an important step for advancing the development of therapeutic agents which modulate, promote or interfere with the differentiation of stem cells.

Hematopoietic stem cells derive from bone marrow stem cells. The bone marrow stem cells ultimately differentiate into the hematopoietic stem cells, which are
15 responsible for the lymphoid, myeloid and erythroid lineages, and stromal stem cells, which differentiate into fibroblasts, osteoblasts, smooth muscle cells, stromal cells and adipocytes (STEWART SELL, IMMUNOLOGY, IMMUNOPATHOLOGY & IMMUNITY, 5th ed. 39-42 Stamford, CT, 1996). The lymphoid lineage, comprising B-cells and T-cells, provides for the production of antibodies, regulation of the cellular immune system, detection of
20 foreign agents in the blood, detection of cells foreign to the host, and the like. The myeloid lineage, which includes monocytes, granulocytes, megakaryocytes as well as others cells, monitors for the presence of foreign bodies in the blood stream, provides protection against neoplastic cells, scavenges foreign materials in the blood stream,

-2-

produces platelets and the like. The erythroid lineage provides the red blood cells which act as oxygen carriers.

Hematopoietic stem cells differentiate as a result from their interaction with growth factors such as interleukins (ILs), lymphokines, colony-stimulating factors 5   (CSFs), erythropoietin (epo), and stem cell factor (SCF). Each of these growth factors have multiple actions that are not necessarily limited to the hematopoietic system (ROBERT A. MEYERS, ED., MOLECULAR BIOLOGY AND BIOTECHNOLOGY: A COMPREHENSIVE DESK REFERENCE, 392-6, New York, 1995). Proliferation, differentiation and survival of immature hematopoietic progenitor cells are sustained by 10   hematopoietic growth factors (hemopoietins). These growth factors also influence the survival and function of mature blood cells. The kinetics of hematopoiesis vary depending on cell type, and their life span may be as little as 6-12 hours to as much as months or years. As a result, the daily renewal of certain lymphocyte progenitors may be substantially lower than that of leukocytic progenitors. The most primitive cells, 15   pluripotent stem cells (PSCs), have high self-renewal capacity (Nathan, 818-821; Saito, *Recent trends in research on differentiation of hematopoietic cells and lymphokines*, Hum. Cell. 5(1): 54 (1992)).

Growth factors are responsible for differentiating the hematopoietic stem cell into either the hemocytoblast, which is the progenitor cell of erythrocytes, neutrophils, 20   eosinophils, basophils, monocytes and platelets, and lymphoid stem cells, which are progenitors to T cells and B cells. SELL, 41. These circulating blood cells are products of terminal differentiation of recognizable precursors (*e.g.*, erythroblasts, mono-myeloblasts and megakaryoblasts, to name but a few). The terminal differentiation of these recognizable precursors may occur exclusively in the marrow cavities of the axial 25   skeleton, with some extension into the proximal femora and humeri (David G. Nathan, *Hematologic Diseases*, IN CECIL TEXTBOOK OF MEDICINE 20th ed., 817, Philadelphia, 1996). White blood cell (WBC) nomenclature may be divided into two major populations on the basis of the form of their nuclei: single nuclei (mononuclear or "round cells") or segmented nuclei (polymorphonuclear).

In human medicine, the ability to initiate and regulate hematopoiesis is of great importance (McCune *et al.,The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function*, Science 241: 1632(1988)). A variety of diseases and immune disorders, including malignancies, appear to be related to

5   disruptions within the lympho-hematopoietic system. Many of these disorders could be alleviated and/or cured by repopulating the hematopoietic system with progenitor cells, which when triggered to differentiate would overcome the patient's deficiency. In humans, a current replacement therapy is bone marrow transplantation. This type of therapy, however, is both painful (for donor and recipient) because of involvement of

10   invasive procedures and can offer severe complications to the recipient, particularly when the graft is allogeneic and Graft Versus Host Disease (GVHD) results. Therefore, the risk of GVHD restricts the use of bone marrow transplantation to patients with otherwise fatal diseases. A potentially more exciting alternative therapy for hematopoietic disorders is the treatment of patients with reagents that regulate the proliferation and

15   differentiation of stem cells (Lawman *et al.*, U.S. Patent No. 5,650,299 (1997)).

There is also a strong interest in the development of procedures to produce large numbers of the human hematopoietic stem cell. This will allow for identification of growth factors associated with its self regeneration. Additionally, there may be as yet undiscovered growth factors associated (1) with the early steps of dedication of the stem

20   cell to a particular lineage; (2) the prevention of such dedication; and (3) the negative control of stem cell proliferation. Availability of large numbers of stem cells would be extremely useful in bone marrow transplantation, as well as transplantation of other organs in association with the transplantation of bone marrow.

An *in vitro* system that permits determination of what agents induce

25   differentiation or proliferation of progenitor cells within a hematopoietic cell population would have many applications. For example, controlled production of red blood cells would permit the *in vitro* production of red blood cell units for clinical replacement (transfusion) therapy. As is well known, transfused red cells are used in the treatment of anemia following elective surgery, in cases of traumatic blood loss, and in the supportive

30   care of, *e.g.*, cancer patients. Similarly, controlled production of platelets would permit

-4-

the *in vitro* production of platelets for platelet transfusion therapy, which may be used in

cancer patients with thrombocytopenia caused by chemotherapy. For both red cells and

platelets, current volunteer donor pools are accompanied by the risk of infectious

contamination, and availability of an adequate supply can be limited. Determination of

5   such compounds would lend itself to developing methods of controlled *in vitro*

production of specified lineage of mature blood cells to circumvent these problems

(Palsson *et al.*, U.S. Patent No. 5,635,386 (1997)).

        Alternatively, agents could be isolated that selectively deplete a particular lineage

of cells from within a hematopoietic cell population and can similarly confer important

10   advantages. For example, production of stem cells and myeloid cells while selectively

depleting T-cells from a bone marrow cell population could be very important for the

management of patients with human immunodeficiency virus (HIV) infection. Since the

major reservoir of HIV is the pool of mature T-cells, selective eradication of the mature

T-cells from a hematopoietic cell mass collected from a patient has considerable potential

15   therapeutic benefit. If one could selectively remove all the mature T-cells from within an

HIV infected bone marrow cell population while maintaining viable stem cells, the T-cell

depleted bone marrow sample could then be used to "rescue" the patient following

hematolymphoid ablation and autologous bone marrow transplantation. Although there

are reports of the isolation of progenitor cells (see, *e.g.*, Tsukamoto *et al.*, (1991) as

20   representative) such techniques are distinct from the selective removal of T-cells from a

hematopoietic tissue culture (Palsson *et al.*, U.S. Patent No. 5,635,386 (1997)).

Summary of the Invention

        While the differentiation of stem cells has been the subject of intense study, little

is known about the global transcriptional response of stem cells during cell

25   hematopoiesis. The present inventors have devised an approach to systematically assess

the transcriptional regulation of stem cells during hematopoiesis as well as methods for

the identification of agents that modulate the expression of at least one gene associated

with hematopoiesis.

-5-

The present invention includes a method to identify stem cell genes that are differentially expressed in stem cells at various stages of differentiation when compared to undifferentiated stem cells by preparing a gene expression profile of a stem cell population and comparing the profile to a profile prepared from stem cells at different
5   stages of differentiation, thereby identifying cDNA species, and therefore genes, which are expressed.

The present invention further includes a method to identify an agent that modulates the expression of at least one stem cell gene associated with the differentiation process of a stem cell population, comprising the steps of preparing a first gene
10   expression profile of an undifferentiated stem cell population, preparing a second gene expression profile of a stem cell population at a defined stage of differentiation, treating said undifferentiated stem cell population with the agent, preparing a third gene expression profile of the treated stem cell population, and comparing the first, second and third gene expression profiles. Comparison of the three gene expression profiles for
15   RNA species as represented by cDNA fragments that are differentially expressed upon addition of the agent to the undifferentiated stem cell population identifies agents that modulate the expression of at least one gene in undifferentiated stem cells that is associated with stem cell differentiation.

Another aspect of the invention is a composition comprising a grouping of nucleic
20   acids or nucleic acid fragments affixed to a solid support. The nucleic acids affixed to the solid support correspond to one or more genes whose expression levels are modulated during stem cell differentiation.

Brief Description of the Drawings

Fig. 1   Figure 1 is an autoradiogram of the gene expression profiles generated
25   from cDNAs made with RNA isolated from Lin[+], LRH, LRH48 and LRBRH cells. All possible 12 anchoring oligo d(T)n1, n2 were used to generate a complete expression profile for the enzyme *ClaI*.

-6-

Modes of Carrying Out the Invention

General Description

The differentiation of stem cells during the process of hematopoiesis is a subject of primary importance in view of the need to find ways to modulate the stem cell

5   differentiation process. One means of characterizing the process of hematopoiesis is to measure the ability of stem cells to synthesize specific RNA during stem cell differentiation.

The following discussion presents a general description of the invention as well definitions for certain terms used herein.

10 *Definitions*

The term "stem cells" as used herein, refers to both hematopoietic stem cells and bone marrow stem cells, and includes totipotent cells which serve as progenitors of neoplastic transformation. The term "hematopoietic stem cells" refers to stem cells which differentiate into erythrocytes, monocytes, granulocytes, and platelets. The

15 putative human hematopoietic stem cell may express the cell surface antigen CD34.

The term "hematopoiesis " as used herein, refers to the process by which stem cells differentiate into blood cells, including erythrocytes, monocytes, granulocytes, and platelets.

The term "blood cell", as used herein, refers to all blood cell types derived from the

20 process of hematopoiesis (see STEWART SELL, *IMMUNOLOGY, IMMUNOPATHOLOGY &* *IMMUNITY*, 5th ed. 39-42, Stamford, CT, 1996)

The term "solid support", as used herein, refers to any support to which nucleic acids can be bound or immobilized, including nitrocellulose, nylon, glass, other solid supports which are positively charged and nanochannel glass arrays disclosed by Beattie (WO

25 95/1175).

The term "gene expression profile", also referred to as a "differential expression profile" or "expression profile" refers to any representation of the expression level of at

-7-

least one mRNA species in a cell sample or population. For instance, a gene expression
profile can refer to an autoradiograph of labeled cDNA fragments produced from total
cellular mRNA separated on the basis of size by known procedures. Such procedures
include slab gel electrophoresis, capillary gene electrophoresis, high performance liquid
5   chromatography, and the like. Digitized representations of scanned electrophoresis gels
are also included as are two and three dimensional representations of the digitized data.

While a gene expression profile encompasses a representation of the expression level
of at least one mRNA species, in practice, the typical gene expression profile represents
the expression level of multiple mRNA species. For instance, a gene expression profile
10   useful in the methods and compositions disclosed herein represents the expression levels
of at least about 5, 10, 20, 50, 100 , 150, 200, 300, 500, 1000 or more preferably,
substantially all of the detectable mRNA species in a cell sample or population.
Particularly preferred are gene expression profiles or arrays affixed to a solid support that
contain a sufficient representative number of mRNA species whose expression levels are
15   modulated under the relevant infection, disease, screening, treatment or other
experimental conditions. In some instances a sufficient representative number of such
mRNA species will be about 1, 2, 5, 10, 15, 20, 25, 30, 40, 50, 50-75 or 100.

Gene expression profiles can be produced by any means known in the art, including,
but not limited to the methods disclosed by: Prashar et al. (1996) *Proc. Natl. Acad. Sci.*
20   *USA* 93:659-663; Liang et al. (1992) *Science* 257:967-971; Ivanova et al. (1995) *Nucleic
Acids Res.* 23:2954-2958; Guilfoyl et al. (1997) *Nucleic Acids Res.* 25(9):1854-1858;
Chee et al. (1996) *Science* 274:610-614; Velculescu et al. (1995) *Science* 270:484-487;
Fischer et al. (1995) *Proc. Natl. Acad. Sci. USA* 92(12):5331-5335; and Kato (1995)
Nucleic Acids Res. 23(18):3685-3690.

25   As an example, gene expression profiles are made to identify one or more genes
whose expression levels are modulated during the process of stem cell differentiation.
The assaying of the modulation of gene expression via the production of a gene
expression profile generally involves the production of cDNA from polyA$^+$ RNA
(mRNA) isolated from stem cells as described below.

-8-

Stem cells are harvested or isolated by any technique known in the art. One of the
most versatile ways to separate hematopoietic cells is by use of flow cytometry, where
the particles, *i.e,*. cells, can be detected by fluorescence or light scattering. The source of
the cells may be any source which is convenient. Thus, various tissues, organs, fluids, or

5   the like may be the source of the cellular mixtures. Of particular interest are bone
marrow and peripheral blood, although other lymphoid tissues are also of interest, such as
spleen, thymus, and lymph node (see Sasaki *et al.,* U.S. Patent No. 5,466,572 and Fei *et
al.,* U.S. Patent No. 5,635,387).

Cells of interest will usually be detected and separated by virtue of surface membrane

10   proteins which are characteristic of the cells. For example, CD34 is a marker for
immature hematopoietic cells. Markers for dedicated cells may include CD 10, CD19,
CD20, and sIg for B cells, CD 15 for granulocytes, CD 16 and CD33 for myeloid cells,
CD 14 for monocytes, CD41 for megakaryocytes, CD38 for lineage dedicated cells, CD3,
CD4, CD7, CD8 and T cell receptor (TCR) for T cells, Thy-1 for progenitor cells,

15   glycophorin for erythroid progenitors and CD71 for activated T cells. In isolating early
progenitors, one may divide a CD34 positive enriched fraction into lineage (Lin)
negative, e.g. CD2 - , CD 14 - , CD15 - , CD16 - , CD10 - , CD19 - , CD33 - and
glycophorin A - , fractions by negatively selecting for markers expressed on lineage
committed cells, Thy-1 positive fractions, or into CD38 negative fractions to provide a

20   composition substantially enriched for early progenitor cells. Other markers of interest
include V alpha and V beta chains of the T-cell receptor (Sasaki *et al.,* U. S. Patent No.
5,466,572 (1995)).

After isolation of the appropriate stem cells, total cellular mRNA is isolated from the
cell sample. mRNAs are isolated from cells by any one of a variety of techniques.

25   Numerous techniques are well known (*see e..,* Sambrook et al., *Molecular Cloning: A
Laboratory Approach*, Cold Spring harbor Press, NY, 1987; Ausbel et., *Current
Protocols in Molecular Biology*, Greene Publishing Co. NY, 1995). In general, these
techniques first lyse the cells and then enrich for or purify RNA. In one such protocol,
cells are lysed in a Tris-buffered solution containing SDS. The lysate is extracted with

30   phenol/chloroform, and nucleic acids precipitated. The mRNAs may be purified from

crude preparations of nucleic acids or from total RNA by chromatography, such as binding and elution from oligo(dT)-cellulose or poly(U)-Sepharose®. However, purification of poly(A)-containing RNA is not a requirement. As stated above, other protocols and methods for isolation of RNAs may be substituted.

5 The mRNAs are reverse transcribed using an RNA-directed DNA polymerase, such as reverse transcriptase isolated from AMV, MoMuLV or recombinantly produced. Many commercial sources of enzyme are available (e.g. Pharmacia, New England Biolabs, Stratagene Cloning Systems). Suitable buffers., cofactors, and conditions are well known and supplied by manufacturers (see also, Sambrook et al. (1989) *Molecular Cloning: a*

10 *laboratory manual*, 2nd Ed., Cold Spring Harbor Laboratory; and Ausbel et al., (1987) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, N.Y.).

 Various oligonucleotides are used in the production of cDNA. In particular, the methods utilize oligonucleotide primers for cDNA synthesis, adapters, and primers for

15 amplification. Oligonucleotides are generally synthesized so single strands by standard chemistry techniques, including automated synthesis. Oligonucleotides are subsequently de-protected and may be purified by precipitation with ethanol, chromatographed using a sized or reversed-phase column, denaturing polyacrylamide gel electrophoresis, high-pressure liquid chromatography (HPLC), or other suitable method. In addition, within

20 certain preferred embodiments, a functional group, such as biotin, is incorporated preferably at the 5' or 3' terminal nucleotide. A biotinylated oligonucleotide may be synthesized using pre-coupled nucleotides, or alternatively, biotin may be conjugated to the oligonucleotide using standard chemical reactions. Other functional groups, such as florescent dyes, radioactive molecules, digoxigenin, and the like, may also be

25 incorporated.

 Partially-double stranded adaptors are formed from single stranded oligonucleotides by annealing complementary single-stranded oligonucleotides that are chemically synthesized or by enzymatic synthesis. Following synthesis of each strand, the two oligonucleotide strands are mixed together in a buffered salt solution (e.g., 1 M NaCl,

30 100 mM Tris-HCl pH.8.0, 10 mM EDTA) or in a buffered solution containing Mg^{+2} (e.g.,

-10-

10 mM MgCl₂) and annealed by heating to high temperature and slow cooling to room

temperature.

The oligonucleotide primer that primes first strand DNA synthesis may comprise a 5'

sequence incapable of hybridizing to a polyA tail of the mRNAs, and a 3' sequence that

5   hybridizes to a portion of the polyA tail of the mRNAs and at least one non-polyA

nucleotide immediately upstream of the polyA tail. The 5' sequence is preferably a

sufficient length that can serve as a primer for amplification. The 5' sequence also

preferably has an average G+C content and does not contain large palindromic sequence;

some palindromes, such as a recognition sequence for a restriction enzyme, may be

10   acceptable. Examples of suitable 5' sequences are CTCTCAAGGATCTACCGCT (SEQ

ID No. _____), CAGGGTAGACGACGCTACGC (SEQ ID No. _____), and

TAATACCGCGCCACATAGCA (SEQ ID No. _____)

The 5' sequence is joined to a 3' sequence comprising sequence that hybridizes to a

portion of the polyA tail of mRNAs and at least one non-polyA nucleotide immediately

15   upstream. Although the polyA-hybridizing sequence is typically a homopolymer of dT

or dU, it need only contain a sufficient number of dT or dU bases to hybridize to polyA

under the conditions employed. Both oligo-dT and oligo-dU primers have been used and

give comparable results. Thus, other bases may be interspersed or concentrated, as long

as hybridization is not impeded. Typically, 12 to 18 bases or 12 to 30 bases of dT or dU

20   will be used. However, as one skilled in the art appreciates, the length need only be

sufficient to obtain hybridization. The non-poly A⁺ nucleotide is A, C, or G, or a

nucleotide derivative, such as inosinate. If one non-polyA nucleotide is used, then three

oligonucleotide primers are needed to hybridize to all mRNAs. If two non-polyA

nucleotides are used, then 12 primers are needed to hybridize to all mRNAs (AA, AC,

25   AG, AT, CA, CC, CG, CT, GA, GC, GG, GT). If three non-poly A nucleotides are used

then 48 primers are needed (3 X 4 X 4). Although there is no theoretical upper limit on

the number of non-polyA nucleotides, practical considerations make the use of one or

two non-polyA nucleotides preferable.

For cDNA synthesis, the mRNAs are either subdivided into three (if one non-polyA

30   nucleotide is used) or 12 (if two non-polyA nucleotides are used) fractions, each

-11-

containing a single oligonucleotide primer, or the primers may be pooled and contacted with a mRNA preparation. Other subdivisions may alternatively be used. Briefly, first strand cDNA is initiated from the oligonucleotide primer by reverse transcriptase (RTase). As noted above, RASE may be obtained from numerous sources and protocols

5   are well known. Second strand synthesis may be performed by RASE (Gubler and Hoffman, *Gene* 25: 263, 1983), which also has a DNA-directed DNA polymerase activity, with or without a specific primer, by DNA polymerase 1 in conjunction with RNaseH and DNA ligase, or other equivalent methods. The double-stranded cDNA is generally treated by phenol:chloroform extraction and ethanol precipitation to remove

10   protein and free nucleotides.

Double-stranded cDNA is subsequently digested with an agent that cleaves in a sequence-specific manner. Such cleaving agents include restriction enzymes, chemical cleaving agents, triple helix, and any other cleaving agent available. Restriction enzyme digestion is preferred; enzymes that are relatively infrequent cutters (*e.g.*, ≥ 5 bp

15   recognition site) are preferred and those that leave overhanging ends are especially preferred. A restriction enzyme with a six base pair recognition site cuts approximately 8% of cDNAs, so that approximately 12 such restriction enzymes should be needed to digest every cDNA at least once. By using 30 restriction enzymes, digestion of every cDNA is assured.

20   The adapters for use in the present invention are designed such that the two strands are only partially complementary and only one of the nucleic acid strands that the adapter is ligated to can be amplified. Thus, the adapter is partially double-stranded (*i.e.*, comprising two partially hybridized nucleic acid strands), wherein portions of the two strands are non-complementary to each other and portions of the two strands are

25   complementary to each other. Conceptually, the adapter may be "Y-shaped" or "bubble-shaped." When the 5' region is non-paired, the 3' end of other strand cannot be extended by a polymerase to make a complementary copy. The ligated adapter can also be blocked at the 3' end to eliminate extension during subsequent amplifications. Blocking groups include dideoxynucleotides and other available blocking agents. In this type of adapter

30   ("Y-shaped"), the non-complementary portion of the upper strand of the adapters is

-12-

preferably a length that can serve as a primer for amplification. As noted above, the non-complementary portion of the lower strand need only be one base, however, a longer sequence is preferable (*e.g.*, 3 to 20 bases; 3 to 15 bases; 5 to 15 bases, or 14 to 24 bases. The complementary portion of the adapter should be long enough to form a duplex under
5   conditions of ligation.

For "bubble-shaped" adapters, the non-complementary portion of the upper strands is preferably a length that can serve as a primer for amplification. Thus, this portion is preferably 15 to 30 bases. Alternatively, the adapter can have a structure similar to the Y-shaped adapter, but has a 3' end that contains a moiety that a DNA polymerase cannot
10 extend from.

Amplification primers are also used in the present invention. Two different amplification steps are performed in the preferred aspect. In the first, the 3' end (referenced to mRNA) of double stranded cDNA that has been cleaved and ligated with an adapter is amplified. For this amplification, either a single primer or a primer pair is
15 used. The sequence of the single primer comprises at least a portion of the 5' sequence of the oligonucleotide primer used for first strand cDNA synthesis. The portion need only be long enough to serve as an amplification primer. The primer pair consists of a first primer whose sequence comprises at least a portion of the 5' sequence of the oligonucleotide primer as described above; and a second primer whose sequence
20 comprises at least a portion of the sequence of one strand of the adapter in the non-complementary portion. The primer will generally contain all the sequence of the non-complementary potion, but may contain less of the sequence, especially when the non-complementary portion is very long, or more of the sequence, especially when the non-complementary portion is very short. In some embodiments, the primer will contain
25 sequence of the complementary portion, as long as that sequence does not appreciably hybridize to the other strand of the adapter under the amplification conditions employed. For example, in one embodiment, the primer sequence comprises four bases of the complementary region to yield a 19 base primer, and amplification cycles are performed at 56°C (annealing temperature), 72°C (extension temperature), and 94°C (denaturation
30 temperature). In another embodiment, the primer is 25 bases long and has 10 bases of

sequence in the complementary portion. Amplification cycles for this primer are performed at 68°C (annealing and extension temperature) and 94°C (denaturation temperature). By using these longer primers, the specificity of priming is increased.

The design of the amplification primers will generally follow well-known guidelines,
5   such as average G-C content, absence of hairpin structures, inability to form primer-dimers and the like. At times, however, it will be recognized that deviations from such guidelines may be appropriate or desirable.

In instances where small numbers of cells are available for the initial RNA extraction, such as small numbers of stem cells, the preferred method of producing a gene expression
10 profile comprises the following general steps. Total RNA is extracted from as few as 5000 stem cells. Using an oligo-dT primer, double stranded cDNA is synthesized and ligated to an adapter in accordance with the present invention. Using adapter primers, the cDNA is PCR amplified using the protocol of Baskaran and Weissman (1996) *Genome-Research* 6(7): 633 and/or Liv *et al.* (1992) *Methods of Enzymology*. The original cDNA
15 is therefore amplified several fold so that a large quantity of this cDNA is available for use in the display protocol according to the present invention. For the display, an aliquot of this cDNA is incubated with an anchored oligo-dT primer. In one method, this mixture is first heat denatured and then allowed to remain at 50°C for 5 minutes to allow the anchor nucleotides of the oligo-dT primers to anneal. This provides for the synthesis
20 of cDNA utilizing Klenow DNA polymerase. The 3'-end region of the parent cDNA (mainly the polyA region) that remains single stranded due to pairing and subsequent synthesis of cDNA by the anchored oligo-dT primer at the beginning of the polyA region, is removed by the 5'-3' exonuclease activity of the T4 DNA polymerase. Following incubation of the cDNA with T4 DNA polymerase for this purpose, dNTPs are added in
25 the reaction mixture so that the T4 DNA polymerase initiates synthesis of the DNA over the anchored oligo-dT primer carrying the heel. The net result of this protocol is that the cDNA with the 3' heel is synthesized for display from the double stranded cDNA as the starting material, rather than RNA as the starting material as occurs in conventional 3'-end cDNA display protocol. The cDNA carrying the 3'-end heel is then subjected to
30 restriction enzyme digestion, ligation, and PCR amplification followed by running the

-14-

PCR amplified 3'-end restriction fragments with the Y-shaped adapter on a display gel. An alternate method is presented in Example 1.

After amplification, the lengths of the amplified fragments are determined. Any procedure that separates nucleic acids on the basis of size and allows detection or

5   identification of the nucleic acids is acceptable. Such procedures include slab gel electrophoresis, capillary gel electrophoresis, 2-dimensional electrophoresis, high performance liquid chromatography, and the like.

Electrophoresis is technique based on the mobility of DNA in an electric field. Negatively charged DNA migrates towards a positive electrode at a rate dependent on

10   their total charge, size, and shape. Most often, DNA is electrophoresed in agarose or polyacrylamide gels. For maximal resolution, polyacrylamide is preferred and for maximal linearity, a denaturant, such as urea is present. A typical gel setup uses a 19:1 mixture of acrylamide:bisacrylamide and a Tris-borate buffer. DNA samples are denatured and applied to the gel, which is usually sandwiched between glass plates. A

15   typical procedure can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Approach*, Cold Spring Harbor Press, NY, 1989) or Ausbel *et al.* (*Current Protocols in Molecular Biology*, Greene Publishing Co., NY, 1995). Variations may be substituted as long as sufficient resolution is obtained.

Capillary electrophoresis (CE) in its various manifestations (free solution,

20   isotachophoresis, isoelectric focusing, polyacrylamide get. micellar electrokinetic "chromatography") allows high resolution separation of very small sample volumes. Briefly, in capillary electrophoresis, a neutral coated capillary, such as a 50 μm X 37 cm column (eCAP neutral, Beckman Instruments, CA), is filled with a linear polyacrylamide (*e.g.*, 0.2% polyacrylamide), a sample is introduced by high-pressure injection followed

25   by an injection of running buffer (*e.g.*, 1X TBE). The sample is electrophoresed and fragments are detected. An order of magnitude increase can be achieved with the use of capillary electrophoresis. Capillaries may be used in parallel for increased throughput (Smith et al. (1990) *Nuc. Acids. Res.* 18:4417; Mathies and Huang (1992) *Nature* 359:167). Because of the small sample volume that can be loaded onto a capillary,

30   sample may be concentrated to increase level of detection. One means of concentration

-15-

is sample stacking (Chien and Burgi (1992) *Anal. Chem* 64:489A). In sample stacking, a large volume of sample in a low concentration buffer is introduced to the capillary column. The capillary is then filled with a buffer of the same composition, but at higher concentration, such that when the sample ions reach the capillary buffer with a lower

5   electric field, they stack into a concentrated zone. Sample stacking can increase detection by one to three orders of magnitude. Other methods of concentration, such as isotachophoresis, may also be used.

High-performance liquid chromatography (HPLC) is a chromatographic separation technique that separates compounds in solution. HPLC instruments consist of a reservoir

10 of mobile phase, a pump, an injector, a separation column, and a detector. Compounds are separated by injecting an aliquot of the sample mixture onto the column. The different components in the mixture pass through the column at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase. IP-RO-HPLC on non-porous PS/DVB particles with chemically

15 bonded alkyl chains can also be used to analyze nucleic acid molecules on the basis of size (Huber et al. (1993) *Anal. Biochem.* 121:351; Huber et al. (1993) *Nuc. Acids Res.* 21:1061; Huber et al. (1993) *Biotechniques* 16:898).

In each of these analysis techniques, the amplified fragments are detected. A variety of labels can be used to assist in detection. Such labels include, but are not limited to,

20 radioactive molecules (*e.g.*, ^{35}S, ^{32}P, ^{33}P), fluorescent molecules, and mass spectrometric tags. The labels may be attached to the oligonucleotide primers or to nucleotides that are incorporated during DNA synthesis, including amplification.

Radioactive nucleotides may be obtained from commercial sources; radioactive primers may be readily generated by transfer of label from γ-^{32}P-ATP to a 5'-OH group

25 by a kinase (*e.g.*, T4 polynucleotide kinase). Detection systems include autoradiograph, phosphor image analysis and the like.

Fluorescent nucleotides may be obtained from commercial sources (*e.g.*, ABI, Foster city, CA) or generated by chemical reaction using appropriately derivatized dyes. Oligonucleotide primers can be labeled, for example, using succinimidyl esters to

30 conjugate to amine-modified oligonucleotides. A variety of florescent dyes may be used,

-16-

including 6 carboxyfluorescein, other carboxyfluorescein derivatives, carboxyrhodamine derivatives, Texas red derivatives, and the like. Detection systems include photomultiplier tubes with appropriate wave-length filters for the dyes used. DNA sequence analysis systems, such as produced by ABI (Foster City, CA), may be used.

5   After separation of the amplified cDNA fragments, cDNA fragments which correspond to differentially expressed mRNA species are isolated, reamplified and sequenced according to standard procedures. For instance, bands corresponding the cDNA fragments can be cut from the electrophoresis gel, reamplified and subcloned into any available vector, including pCRscript using the PCR script cloning kit (Stratagene).

10   The insert is then sequenced using standard procedures, such as cycle sequencing on an ABI sequencer (Foster City, CA).

An additional means of analysis comprises hybridization of the amplified fragments to one or more sets of oligonucleotides immobilized on a solid substrate. Historically, the solid substrate is a membrane, such as nitrocellulose or nylon. More recently, the

15   substrate is a silicon wafer or a borosilicate slide. The substrate may be porous (Beattie et al. WO 95/11755) or solid. Oligonucleotides are synthesized in situ or synthesized prior to deposition on the substrate using standard procedures. Various chemistries are known for attaching oligonucleotides. Many of these attachment chemistries rely upon functionalizing oligonucleotides to contain a primary amine group. The oligonucleotides

20   are arranged in an array form, such that the position of each oligonucleotide sequence can be determined.

The amplified fragments, which are generally labeled according to one of the methods described herein, are denatured and applied to the oligonucleotides on the substrate under appropriate salt and temperature conditions. In certain embodiments, the conditions are

25   chosen to favor hybridization of exact complementary matches and disfavor hybridization of mismatches. Unhybridized nucleic acids are washed off and the hybridized molecules detected, generally both for position and quantity. The detection method will depend upon the label used. Radioactive labels, fluorescent labels and mass spectrometry label are among the suitable labels.

-17-

The present invention as set forth in the specific embodiments, includes methods to identify a therapeutic agent that modulates the expression of at least one stem cell gene associated with the differentiation, proliferation and/or survival of stem cells.

As an example, the method to identify an agent that modulates the expression of at

5   least one stem cell gene associated with the differentiation of a stem cell population, comprises the steps of preparing a first gene expression profile of an undifferentiated stem cell population, preparing a second gene expression profile of a stem cell population at a defined stage of differentiation, treating said undifferentiated stem cell population with the agent, preparing a third gene expression profile of the treated stem cell

10   population, and comparing the first, second and third gene expression profiles. Comparison of the three gene expression profiles for RNA species as represented by cDNA fragments that are differentially expressed upon addition of the agent to the undifferentiated stem cell population identifies agents that modulate the expression of at least one gene in undifferentiated stem cells that is associated with stem cell

15   differentiation.

While the above methods for identifying a therapeutic agent comprise the comparison of gene expression profiles from treated and not-treated stem cells, many other variations are immediately envisioned by one of ordinary skill in the art. As an example, as a variation of a method to identify a therapeutic agent that modulates the expression of at

20   least one stem cell gene associated with the differentiation, the second gene expression profile of a stem cell population at a defined stage of differentiation and the third gene expression profile of the treated stem cell population can each be independently normalized using the first gene expression profile prepared from the undifferentiated stem cell population. Normalization of the profiles can easily be achieved by scanning

25   autoradiographs corresponding to each profile, and subtracting the digitized values corresponding to each band on the autoradiograph from undifferentiated stem cells from the digitized value for each corresponding band on autoradiographs corresponding to the second and third gene expression profiles. After normalization, the second and third gene expression profiles can be compared directly to detect cDNA fragments which

-18-

correspond to mRNA species which are specifically expressed during differentiation of a stem cell population.

Specific Embodiments

Example 1

5   *Production of gene expression profiles generated from cDNAs made with RNA isolated from undifferentiated and partially differentiated stem cells.*

Crude Marrow Preparation

Expression profiles of RNA expression levels from undifferentiated stem cells and stems cells at various levels of differentiation, including partially differentiated and

10 terminally differentiated stem cells, offer a powerful means of identifying genes whose expression levels are associated with stem cell differentiation or proliferation. As an example, the production of expression profiles from murine lineage negative, rhodamine low, Hoechst low and rhodamine bright, Hoechst low hematopoietic precursor cells allows for the identification of mRNA species and their encoding genes whose

15 expression levels are associated with stem cell differentiation

Hoechst^{low}/Rhodamine^{low} hematopoietic stem cells were isolated by sacrificing 30 Balb/c female mice (6-12 weeks) and surgically removing the iliac crests, femurs and tibiae. The bones were cleaned and placed in 10 ml PBS/5% HI-FBS on ice. One tube was used for the bones from 10 mice. The bones were ground throughly with a pestle

20 until completely broken. Following grinding, the supernatant was removed into a 50 ml conical tube through a 40 μM filer(Falcon #2340). 10 ml PBS/FBS was added to the mix and the supernatant removed. The supernatant was then centrifuged (1250 rpm) for 5-10 minutes. The supernatant which contains a high concentration of lipid was then decanted and discarded.

25   The cells were then pooled into 25 or 50 ml fresh PBS/FBS, and tiny bone fragments removed by settling. The cells were then counted in crystal violet. Cells were diluted and underlayed with LSM, centrifuged at 2000rpm(1000xg) for 20 minutes. To harvest the buffy coat, the supernatant was removed to within 1 cm of the cells. The next 8-

-19-

10ml of medium and cells were harvested by swirling the media around in the tube to draw cells from all sides of the gradient. The cell volume was then brought up to 50 ml with PBS/FBS and spun at 1400rpm 5-10 minutes.

Lineage Depletion

5   Cells were counted in Crystal Violet and resuspended in fresh PBS/FBS. Lineage-specific antibodies were added as follows:

TER 119	$0.1\mu g/ml$ final concentration
B220	$15\mu l/10^8$ cells
Mac-1	$15\mu l/10^8$ cells
Gr-1	$15\mu l/10^8$ cells
Lyt-2	1/20 final dilution
L3T4	1/20 final dilution
Yw25.12.7	1/100 final dilution

The cells were incubated on ice for 15 minutes, brought to a volume of 50ml with

15   PBS/FBS and collected at 1400rpm for 5-10 minutes, and washed to remove unbound antibodies.

During the antibody binding step, Magnetic Beads(Dynabeads M-450) were prepared at a ratio of 5 beads/cell. The beads were coated with Sheep anti-Rat antibodies that bind to the lineage-specific antibodies, which are all of rat origin. When the beads are placed in

20   a magnetic field, the Lin^+ cells are removed. The resulting supernatant contains the Lin^- population (granulocytes and lymphocyte populations will be substantially depleted or absent after this step.)

Hoechst/Rhodamine Staining

Rhodamine 123 was added to a final concentration of 0.1 μg/ml, then incubated at

25   32°C for 20 minutes in the dark. Without further manipulation or washing, HOECHST 33342 was added to a final concentration of 10μM then incubated at 37°C for an additional hour. The aliquot of crude marrow was brought to 0.5 ml with PBS/FBS and Hoechst to this cell preparation as well. The volume was brought to 50 ml with PBS/FBS, centrifuged at 1400rpm for 5-10 minutes, supernatant discarded and cells

30   resuspended to $2x10^7$ cells/ml. The rhodamine only and Hoechst Only/Crude Marrow

-20-

were washed in parallel. These two populations were then resuspended in 0.5ml

PBS/FBS for flow cytometry analysis

   Total RNA was extracted from approximately 5000 stem cells. Using an oligo-dT

primer, double stranded cDNA is synthesized and ligated to an adapter in accordance

5   with the present invention. Using adapter primers, the cDNA is PCR amplified using the

protocol of Baskaran and Weissman (1996) *Genome Research* 6(7): 633 and Lie *et al.,*

Methods of Enzymology, ____ . The original cDNA is therefore amplified several fold so

that a large quantity of this cDNA is available for use in the display protocol according to

the present invention.

10   Synthesis of cDNA for the gene expression profiles was performed as below:

Materials and Reagents

   A microPoly(A)Pure mRNA Isolation kit (Ambion Inc.) was used for mRNA isolation.

All the reagents for cDNA synthesis were obtained from Life Technologies Inc. Klentaq1

DNA polymerase ($25U/\mu l$) was from Ab peptides Inc. Native *Pfu* DNA polymerase

15   ($2.5U/\mu l$) was purchased from Stratagene Inc. Betaine monohydrate was from Fluka

BioChemica and dimethylsulfoxide (DMSO) was from Sigma Chemical Company.

Deoxynucleoside triphophates (dNTPs, 100mM) and bovine serum albumin (BSA, 10

mg/ml) were purchased from New England Biolabs, Inc. Qiaquick PCR purification kit

(Qiagen) was used to purify the amplified PCR products. The oligonucleotides used in the

20   Examples were synthesized and gel purified in the DNA synthesis laboratory (Department

of Pathology, Yale University School of Medicine, New Haven, CT).

Table 1. Sequences of oligonucleotides.

T_7-SalI-oligo-d(T)V	5'-ACG TAA TAC GAC TCA CTA TAG GGC GAA TTG GGT CGA C-d (T)$_{18}$ V-3' , where V = A, C, G
anti-NotI Long	5'-CTT ACA GCG GCC GCT TGG ACG-3'

-21-

NotI Short	5'-AGC GGC CGC TGT AAG-3'
NotI/RI primer	5'-GCG GAA TTC CGT CCA AGC GGC CGC TGT AAG-3'

Methods

I. Preparation of mRNA

5   MicroPoly(A)Pure mRNA isolation kit was used for the isolation of Poly(A)$^+$ RNA
following the kit instructions. mRNA from a small number of mouse hematopoietic cells
(5,000-10,000 cells) was extracted, eluted from the column, and precipitated by adding 0.1
volume of 5M ammonium acetate and 2.5 volumes of chilled ethanol with 2μg glycogen as
carrier. The tubes were left at -20°C overnight. The pellets were collected by centrifugation
10   at top speed for 30 minutes, washed with 70% ethanol and air-dried at room temperature.
The pellets were resuspended in 10μl H$_2$O/0.1mM EDTA solution. We observed that the
dissolved mRNA solution was cloudy due to the leaching of column materials, therefore the
samples were centrifuged at 4°C for 5 minutes. The supernatant was collected for further
use.

15   II. cDNA synthesis
       First strand cDNA synthesis
       The cDNA synthesis reaction (final reaction volume is 20μl) was carried out as
described in the instruction manual (Superscript Choice System) provided by Life
Technologies Inc. For the first strand cDNA synthesis, mRNA (10μl) isolated from a small
20   number of cells was annealed with 200ng (1μl) of T$_7$-SalI-oligo-d(T)V-primer (see Table-1)
in a 0.5-ml micro centrifuge tube (no stick, USA Scientific Plastics) by heating the tubes at
65°C for 5 minutes, followed by quick chilling on ice for 5 minutes. This step was repeated

-22-

once and the contents were collected at the bottom of the tube by a brief centrifugation. The following components were added to the primer annealed mRNA on ice prior to initiating the reaction, 1μl of 10mM dNTPs, 4μl of 5 x first strand buffer [250mM Tris-HCl (pH 8.3), 375mM KCl, 15mM MgCl$_2$], 2μl of 100mM DTT and 1μl of RNase Inhibitor (40U/μl). All

5   the contents were mixed gently and the tubes were pre-warmed at 45⁰C for 2 minutes. The cDNA synthesis was initiated by adding 200 units (1μl) of Superscript II Reverse Transcriptase and the incubation continued at 45°C for 1 hour.

Second strand cDNA synthesis

At the end of first strand cDNA synthesis, the tubes were kept on ice. Second

10   strand cDNA synthesis reaction (final volume is 150μl) was set up in the same tube on ice by adding 91μl of nuclease free water, 30μl of 5x second strand buffer [100mM Tris-HCl (pH 6.9), 23mM MgCl$_2$, 450mM KCl, 0.75mM (β-NAD⁺ and 50mM ammonium sulfate], 3μl of 10mM dNTPs, 1μl of *E.coli* DNA ligase (10U/μl), 4μl of *E.coli* DNA polymerase I (l0U/μl) and 1μl of *E.coli* RNase H (2U/μl). The contents were

15   mixed gently and the tubes were incubated at 16°C for 2 hours. Following the incubation, the tubes were kept on ice, 2μl of T$_4$ DNA polymerase (3U/μl) was added and the incubation was continued for another 5 minutes at 16°C. The reaction was stopped by the addition of 10μl of 0.5M EDTA (pH 8.0) and extracted once with equal volume of phenol: chloroform 1:1 (v/v) and once with chloroform. The aqueous phase was then

20   transferred to a new tube and precipitated by adding 0.5 volumes of 7.5M ammonium acetate (pH 7.6), 2μg of glycogen (as carrier) and 2.5 volumes of chilled ethanol. The samples were left at -20⁰C for overnight and the cDNA pellets were collected by centrifugation at top speed for 20 minutes. The pellets were washed once with 70% ethanol, air-dried and dissolved in 14μl of nuclease free water.

25   As the amount of cDNA derived from a small number of cells may be low, it may be necessary to amplify the cDNA for further analysis. To uniformly amplify the cDNA, an adaptor (NotI adaptor) was first ligated to both ends of the cDNA. Following adaptor

-23-

ligation, the cDNAs were amplified with NotI/RI primer (see *table 1*), by a modified
PCR method using betaine and DMSO.

Ligation of cDNA with NotI adaptor

Preparation of NotI adaptor: The NotI adaptor was prepared by annealing
5   NotI-short and anti-NotI-long oligonucleotides (see Table 1). The anti-NotI-long
oligonucleotide was phosphorylated to ensure that both the adaptor oligonucleotides are
ligated to the cDNA. 1μg of anti-NotI-long was mixed with 1μl of 10x T$_4$ polynucleotide
kinase buffer [700mM Tris-HCl (pH 7.6), 100mM MgCl$_2$ and 50mM DTT], 1μl of
10mM adenosine triphosphate (ATP), adjusted the volume to 9μl with water and the
10   reaction was initiated by adding 1μl of T$_4$ polynucleotide kinase (10U/μl). The tubes were
incubated at 37°C for 30 minutes and then the enzyme was inactivated at 65°C for 20
minutes. The annealing was carried out by adding the following components to the above
phosphorylated anti-NotI-long: 1μg of NotI-short, 2μl of 10x oligo annealing buffer
[100mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0) and 1M NaCl] and water to adjust
15   the final volume to 20μl. The sample was heated at 65°C for 10 minutes and allowed to
cool down to room temperature. The annealed adaptor was stored at -20°C.

Ligation of cDNA with annealed NotI adaptor: To set up this reaction,
14μl of cDNA was mixed with 100ng of annealed NotI adaptor in a 0.5-ml micro
centrifuge tube. To this mixture 2μl of 10x T$_4$ DNA ligase buffer [500mM Tris-HCl (pH
20   7.8), 100mM MgCl$_2$, 100mM DDT, 10mM ATP and 250mg/ml BSA] was added and
adjusted the volume with water to 18μl and mixed gently. The reaction was initiated by
adding 2μl of T$_4$ DNA ligase (400U/μl) and incubated at 16°C overnight.

III.   cDNA amplification

A modified betaine-DMSO PCR method (Baskaran *et al.* (1996)) Genome
25   Research 6:633) was used to uniformly amplify the cDNA with different GC content.
This method uses the LA system, which combines a highly thermostable form of *Taq*
DNA polymerase (Klentaq1, which is devoid of 5'-exonuclease activity) and a
proofreading enzyme (*Pfu* DNA polymerase, which has 3'-exonuclease activity). The

-24-

LA16 enzyme consists of 1 part of *Pfu* DNA polymerase and 15 parts of KlenTaq1 DNA
Polymerase (v/v). The NotI adaptor-ligated cDNA was diluted 10 fold with water. 2 μl of
this diluted cDNA was used as the template for PCR. The PCR reaction (50μl final
volume) was set up with the following components: 5μl of 10x PCR buffer [200mM

5   Tris-HCl (pH 9.0), 160mM ammonium sulfate and 25mM MgCl$_2$], 16μl of water, 0.8μl
of BSA (10mg/ml), 1μl of NotI/RI PCR primer (100ng/ul), 5μl of 50% DMSO (v/v), 15μl
of 5M Betaine and 0.2μl of LA16 enzyme. These components were mixed gently on ice
and then heated to 95°C for 15 seconds on a PCR machine, and held at 80°C while 5μl of
2mM dNTPs were added to start the reaction. The PCR conditions were as follows: *Stage*

10   *1:* 95°C for 15 seconds, 55°C for 1 minute, 68°C for 5 minutes, 5 cycles. *Stage 2:* 95°C
for 15 seconds, 60°C for 1 minute, 68°C for 5 minutes, 15 cycles.

After amplification, cDNA was purified with the Qiaquick PCR purification kit
(following the instructions provided by the supplier). The purified cDNA was eluted in
the desired volume of water.

15   Gene expression profiles were prepared from the purified cDNA as previously
described by Prashar *et al.* in WO 97/05286 and in Prashar *et al.* (1996) *Proc. Natl. Acad.
Sci. USA* 93:659-663. Briefly, the adapter oligonucleotide sequences were
CTTACAGCGGCCGCTTGGACG, GAATGTCGCCGGCGA or alternatively,
A1 (TAGCGTCCGGCGCAGCGACGGCCAG) and

20   A2 (GATCCTGGCCGTCGGCTGTCTGTCGGCGC). When A1/A2 were used, one
microgram of oligonucleotide A2 was first phosphorylated at the 5' end using T4
polynucleotide kinase (PNK). After phosphorylation, PNK was heated denatured, and
1μg of the oligonucleotide A1 was added along with 10× annealing buffer (1 M
NaCl/100 mM Tris-HCl, pH8.0/10 mM EDTA, pH8.0) in a final vol of 20 μl. This

25   mixture was then heated at 65°C for 10 min followed by slow cooling to room
temperature for 30 min, resulting in formation of the Y adapter at a final concentration of
100 ng/μl. About 20 ng of the cDNA was digested with 4 units of a restriction enzyme
such as *Cla*I, *Bgl* II, etc. in a final vol of 10 μl for 30 min at 37°C. Two microliters ($\approx$4
ng of digested cDNA) of this reaction mixture was then used for ligation to 100 ng ($\approx$50-

30   fold) of the Y-shaped adapter in a final vol of 5μl for 16 hr at 15°C. After ligation, the

-25-

reaction mixture was diluted with water to a final vol of 80 μl (adapter ligated cDNA concentration, $\approx$50 pg/μl) and heated at 65°C for 10 min to denature T4 DNA ligase, and 2-μl aliquots (with $\approx$100 pg of cDNA) were used for PCR.

The following sets of primers were used for PCR amplification of the adapter

5   ligated 3'-end cDNAs: GCGGAATTCCGTCCAAGCGGCCGCTGTAAG or alternatively, RP 5.0 (CTCTCAAGGATCTTACCGCTT $_{18}$AT), RP 6.0 (TAATACCGCGCCACATAGCAT $_{18}$CG), or RP 9.2 (CAGGGTAGACGACGCTACGCT$_{18}$GA) were used as 3' primer while A1.1 (TAGCGTCCGGCGCAGCGAC) served as the 5' primer. To detect the PCR products

10   on the display gel, 24 pmol of oligonucleotide A1.1 was 5'-end-labeled using 15 μl of [γ-^{32} P]ATP (Amersham; 3000 Ci/mmol) and PNK in a final volume of 20 μl for 30 min at 37°C. After heat denaturing PNK at 65°C for 20 min, the labeled oligonucleotide was diluted to a final concentration of 2 μM in 80 μl with unlabeled oligonucleotide A1.1. The PCR mixture (20μl) consisted of 2 μl ($\approx$100 pg) of the template, 2μl of 10$\times$ PCR

15   buffer (100 mM Tris·HCl, pH 8.3/500 mM KCl), 2 μl of 15 mM MgCl$_2$ to yield 1.5 mM final Mg^{2+} concentration optimum in the reaction mixture, 200 μM dNTPs, 200 nM each 5' and 3' PCR primers, and 1 unit of Amplitaq. Primers and dNTPs were added after preheating the reaction mixture containing the rest of the components at 85°C. This "hot start" PCR was done to avoid artefactual amplification arising out of arbitrary annealing

20   of PCR primers at lower temperature during transition from room temperature to 94°C in the first PCR cycle. PCR consisted of 28-30 cycles of 94°C for 30 sec, 50°C for 2 min, and 72°C for 30 sec. A higher number of cycles resulted in smeary gel patterns. PCR products (2.5μl) were analyzed on 6% polyacrylamide sequencing gel. For double or multiple digestion following adapter ligation, 13.2 μl of the ligated cDNA sample was

25   digested with a secondary restriction enzyme(s) in a final vol of 20 μl. From this solution, 3μl was used as template for PCR. This template vol of 3 μl carried $\approx$ 100 pg of the cDNA and 10 mM MgCl$_2$ (from the 10$\times$ enzyme buffer), which diluted to the optimum of 1.5 mM in the final PCR vol of 20 μl. Since Mg^{2+} comes from the restriction enzyme buffer, it was not included in the reaction mixture when amplifying

30   secondarily cut cDNA. Bands may then be extracted from the display gels as described

-26-

by Liang *et al.* (1995 *Curr. Opin. Immunol.* 7:274-280), reamplified using the 5' and 3'
primers, and subcloned into pCR-Script with high efficiency using the PCR-Script
cloning kit from Stratagene. Plasmids were sequenced by cycle sequencing on an ABI
automated sequencer.

5   Figure 1 presents an autoradiogram of the gene expression profiles generated
from cDNAs made with RNA isolated from Lin⁺, LRH, LRH48 and LRBRH cells. All
possible 12 anchoring oligo d(T)n1, n2 were used to generate a complete expression
profile for the enzyme *Cla*I.

Table 2 presents the sequences of numerous differentially expressed bands from
10   expression profiles made from LIN⁺, LRH, LRH48 and LRBRH.

Table 3 presents the expression patterns of the differentially expressed bands set
forth in Table 2. The band fragment length (size) in Table 3 is the length before
unwanted terminal sequences were removed. Table 3 also presents the results of a
GenBank Search and analysis of the sequences of Table 2.

Summary of Known Genes from Mouse HSC Differential Display (I)

Items No.	Size (bp)	Enzyme	N1N2 (oligo-dT)	Poly(A) Sign	Lin⁻	LRII	LRII4R	LRIIRII	Gene Bank Search & Analysis
HSC-DD-006	213	Bgl II	AC	fair	0	3+	/	+	mouse homeobox protein
HSC-DD-285	158	Xba I	GG	good	±	+	+	±	human homeobox gene regulator
HSC-DD-007B	213	Bgl II	AC	fair	±	2+	/	±	human zinc finger protein 10
HSC-DD-238	363	Xba I	AG	good	3+	0	3+	3+	mouse cell division control protein 19
HSC-DD-206	123	Xba I	AC	good	3+	0	2+	+	human HS1 heamatopoietic protein
HSC-DD-214	192	Xba I	AC	fair	±	2+	0	3+	mouse plm-1 proto-oncogene
HSC-DD-035	151	Bgl II	AC	fair	±	2+	/	+	mouse thyroid hormone receptor
HSC-DD-129	234	Cla I	AC	poor	0	3+	0	0	mouse inositol 1,4,5-trisphosphate receptor
HSC-DD-040	220	Bgl II	AC	fair	+	2+	/	0	mouse G protein beta-36 subunit
HSC-DD-011	173	Bgl II	AC	good	±	±	/	2+	mouse ras-related YPT1 protein
HSC-DD-121	186	Cla I	CT	poor	0	3+	±	±	human TBP-associated factor 170
HSC-DD-015B	133	Bgl II	AG	poor	0	3+	/	+	mouse HMG1-related DNA binding protein
HSC-DD-039	206	Bgl II	AC	fair	2+	4+	/	4+	mouse TAX responsive element binding protein 107
HSC-DD-042	235	Bgl II	AC	fair	±	0	/	+	mouse retinoblastoma binding protein isoform III
HSC-DD-256	272	Xba I	AA	poor	0	2+	±	0	Rat androgen-binding protein
HSC-DD-045	270	Bgl II	AC	good	±	2+	/	±	similar to Rat cea2
HSC-DD-068	164	Cla I	AC	fair	+	4+	4+	4+	mouse jerky mRNA
HSC-DD-143	350	Cla I	AG	fair	±	2+	/	±	similar to human memd
HSC-DD-263	292	Xba I	AT	good	0	2+	±	0	mouse interleukin 5
HSC-DD-239	156	Xba I	CA	good	±	3+	3+	+	human CD9
HSC-DD-261	115	Xba I	AA	good	0	*	0	0	mouse germline IgM
HSC-DD-028A	95	Bgl II	AC	good	*	4+	4+	*	mouse chaperonin containing TCP-1 e subunit
HSC-DD-021	143	Bgl II	AG	low	±	*	/	2*	mouse calreticulin
HSC-DD-025	326	Bgl II	AG	good	±	2+	/	2*	mouse metallothionein I

Summary of Known Genes from Mouse HSC Differential Display (II)

Items No.	Size (bp)	Enzyme	NIN2 (oligo-dT)	Poly(A) Size	Lin*	LRII	LKII4R	LRIIRII	Gene Bank Search & Analysis
HSC-DD-077	203	Cla I	AC	good	+	2+	2+	3+	Rat matrin cyclophilin
HSC-DD-200	450	Cla I	AA	fair	+	±	2+	+	mouse G-utrophin
HSC-DD-245	272	Xba I	CA	fair	3+	±	3+	2+	rat basement membrane-associated chondroitin
HSC-DD-226	387	Xba I	AC	good	±	3+	±	0	mouse cytoplasmic g-actin
HSC-DD-182	149	Cla I	GC	poor	±	3+	±	+	mouse A-X actin
HSC-DD-089	364	Cla I	AC	poor	+	3+	2+	+	mouse TIE receptor tyrosine kinase
HSC-DD-151	424	Cla I	GA	good	0	+	2+	±	rat elk, brain-specific receptor tyrosine kinase
HSC-DD-013	248	Bgl II	AC	fair	±	2+	/	3+	mouse hexokinase
HSC-DD-029	103	Bgl II	AC	fair	0	+	/	0	mouse bruton agammaglobulinemia tyrosine kinase
HSC-DD-034	140	Bgl II	AC	fair	0	2+	/	2+	mouse spermine synthase
HSC-DD-082B	244	Cla I	AC	fair	+	4+	2+	2+	mouse stearoyl-CoA desaturase (SCD2)
HSC-DD-084	261	Cla I	AC	good	±	+	±	2+	mouse antioxidant enzyme AOE 372
HSC-DD-128	189	Cla I	AC	fair	0	3+	3+	±	mouse casein kinase II beta chain
HSC-DD-140	229	Cla I	AG	good	±	0	0	+	mouse creatine kinase B
HSC-DD-148	313	Cla I	GA	good	+	+	2+	±	human esterase D
HSC-DD-176	470	Cla I	CG	fair	±	3+	+	0	mouse putative E1-E2 ATPase
HSC-DD-178	130	Cla I	GC	good	±	3+	0	+	mouse aspartate aminotransferase
HSC-DD-180	142	Cla I	GC	good	+	+	0	+	mouse tyrosylprotein sulfotransferase-1
HSC-DD-186	252	Cla I	GC	poor	±	+	2+	2+	mouse ubiquitin-conjugating enzyme E214K
HSC-DD-191	136	Cla I	AA	fair	0	±	3+	2+	mouse b-1,4-galactosyltransferase
HSC-DD-158	391	Cla I	GT	fair	+	3+	0	+	spermophilus tridecemlineatus 26s proteasome
HSC-DD-099	265	Cla I	CC	fair	±	3+	0	±	mouse proteasome epsilon chain precursor
HSC-DD-222	270	Xba I	AC	good	0	2+	3+	+	Rat 3-hydroxyiso- butyrate
HSC-DD-104	368	Cla I	CC	fair	0	±	+	±	human copper chaperone for superoxide dismutase
HSC-DD-172	365	Cla I	CG	fair	±	3+	2+	0	mouse Ercc-4 DNA repair gene
HSC-DD-169	223	Cla I	CG	fair	±	±	2+	0	Cricetulus griseus nucleotide excision repair protein
HSC-DD-003A	148	Bgl II	AC	poor	0	+	/	±	human G rich sequence factor

Summary of Known Genes from Mouse HSC Differential Display (III)

Items No.	Size (bp)	Enzyme	NIN2 (oligo-dT)	Poly(A) Sign	Expression pattern				Gene Bank Search & Analysis
					Lin+	LRII	LRH48	LRBRII	
HSC-DD-092	118	Cla I	CC	fair	+	3+	±	+	mouse elongation factor 1-a
HSC-DD-288	480	Xba I	GC	fair	±	+	+	±	human elongation factor-1-delta
HSC-DD-114	267	Cla I	CA	poor	±	+	±	+	Rat elongation factor-1-alpha
HSC-DD-213	178	Xba I	AC	fair	±	3+	+	+	human splicing factor (SFRS7)
HSC-DD-155	198	Cla I	GT	fair	0	2+	+	0	mouse transcription elongation factor S-II-T1
HSC-DD-212	162	Xba I	AC	poor	0	3+	±	0	mouse translation initiation factor 4E
HSC-DD-090	375	Cla I	AC	fair	±	3+	3+	+	mouse protein synthesis elongation factor
HSC-DD-173	367	Cla I	CG	fair	±	3+	+	0	mouse protein synthesis elongation factor Tu
HSC-DD-249	304	Xba I	CA	poor	4+	+	4+	4+	rat histone macroH2A1.2
HSC-DD-250	356	Xba I	CA	good	+	2+	3+	2+	mouse MER9 processed pseudogene
HSC-DD-108	281	Cla I	GG	good	+	2+	+	2+	mouse heat shock protein 70
HSC-DD-116	326	Cla I	CA	fair	±	2+	0	2+	mouse 84 kD heat shock protein
HSC-DD-166	587	Cla I	AT	good	±	2+	3+	+	mouse heat shock protein 70 cognate
HSC-DD-184	196	Cla I	GC	fair	±	2+	0	±	mouse breast heat shock protein 73
HSC-DD-101	331	Cla I	CC	fair	+	3+	0	±	mouse MHC locus II region
HSC-DD-017	215	Bgl II	AG	good	0	4+	/	0	mouse MHC class III region
HSC-DD-026	505	Bgl II	AG	fair	2+	4+	/	4+	mouse ribosomal protein S4
HSC-DD-064	146	Cla I	AC	good	2+	2+	2+	3+	mouse ribosomal protein S12
HSC-DD-066	150	Cla I	AC	good	2+	3+	2+	2+	mouse ribosomal protein S20
HSC-DD-041	226	Bgl II	AC	good	+	3+	/	3+	mouse ribosomal protein L7
HSC-DD-111	161	Cla I	CA	low	±	+	±	+	rat ribosomal protein L23a
HSC-DD-0268	100	Bgl II	AC	low	+	4+	/	+	mouse LINE-1/L1 element
HSC-DD-142	267	Cla I	AG	low	±	2+	±	±	mouse 1.9kb A13 repetitive sequence
HSC-DD-095	210	Cla I	CC	low	±	2+	±	±	mouse mitochondrial 12S ribosomal RNA

-30-

As is apparent to one of ordinary skill in the art, this same procedure can be used to identify stem cells genes whose expression levels are associated with stem cell proliferation, dedicated differentiation and survival.

5   Example 2

Method to identify a therapeutic agent that modulates the expression of at least one stem cell gene associated with the differentiation process of a stem cell population.

The methods set forth in Example 1 offer a powerful approach for identifying therapeutic agents that modulate the expression of at least one stem cell gene associated

10   with the differentiation process of a stem cell population. For instance, gene expression profiles of undifferentiated stem cells and partially differentiated or terminally differentiated stem cells are prepared as set forth in Example 1. A profile is also prepared from an undifferentiated stem cell sample that has been exposed to the agent to be tested. By examining for differences in the intensity of individual bands between the three

15   profiles, agents which up or down regulate genes associated with the differentiation process of a stem cell population are identified.

Example 3

Method to identify a therapeutic agent that modulates the expression of at least one stem

20   *cell gene associated with the proliferation of a stem cell population.*

The methods set forth in Example 1 offer a powerful approach for identifying therapeutic agents that modulate the expression of at least one stem cell gene associated with the proliferation of a stem cell population. For instance, gene expression profiles of undifferentiated stem cells and actively proliferating stem cells are prepared as set forth

25   in Example 1. A profile is also prepared from an undifferentiated stem cell sample that has been exposed to the agent to be tested. By examining for differences in the intensity of individual bands between the three profiles, agents which up or down regulate genes associated with the proliferation of a stem cell population are identified.

SUBSTITUTE SHEET (RULE 26)

-31-

As is apparent to one of ordinary skill in the art, this same procedure can be used

to identify stem cells genes whose expression levels are associated with stem cell

dedicated differentiation and survival.

Example 4

5   *Production of solid support compositions comprising groupings of nucleic acids or*

nucleic acid fragments that correspond to genes whose expression levels are associated

with the differentiation, proliferation, dedicated differentiation or survival of stem cells.

As set forth in Example 1, expression profiles prepared from stem cells at

different stages of differentiation, from proliferating stem cells, from stem cells that are

10   dedicated to a differentiation pathway and from stem cells resistant to apoptosis (which

may be linked to increased survival) provide a means to identify genes whose expression

levels are associated with stem cell differentiation, proliferation, dedicated differentiation

and survival, respectively.

Solid supports can be prepared that comprise immobilized representative

15   groupings of nucleic acids or nucleic acid fragments corresponding to the genes from

stem cells whose expression levels are modulated during stem cell differentiation,

proliferation, dedicated differentiation and survival. For instance, representative nucleic

acids can be immobilized to any solid support to which nucleic acids can be immobilized,

such as positively charged nitrocellulose or nylon membranes (see Sambrook *et al.*

20   (1989) *Molecular Cloning: a Laboratory Manual*, 2nd Ed., Cold Spring Harbor

Laboratory) as well as porous glass wafers such as those disclosed by Beattie (WO

95/11755). Nucleic acids are immobilized to the solid support by well established

techniques, including charge interactions as well as attachment of derivatized nucleic

acids to silicon dioxide surfaces such as glass which bears a terminal epoxide moiety. At

25   least one species of nucleic acid molecule, or fragment of a nucleic acid molecule

corresponding to the genes from stem cells whose expression levels are modulated during

stem cell differentiation, proliferation, dedicated differentiation and survival may be

immobilized to the solid support. A solid support comprising a representative grouping

of nucleic acids can then be used in standard hybridization assays to detect the presence

SUBSTITUTE SHEET (RULE 26)

-32-

or quantity of one or more specific nucleic acid species in a sample (such as a total
cellular mRNA sample or cDNA prepared from said mRNA) which hybridize to the
nucleic acids attached to the solid support. Any hybridization methods, reactions,
conditions and/or detection means can be used , such as those disclosed by Sambrook *et*
5   *al.* (1989) *Molecular Cloning: a Laboratory Manual*, 2nd Ed., Cold Spring Harbor
Laboratory, Ausbel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene
Publishing and Wiley-Interscience. N.Y. or Beattie in WO 95/11755.

       One of ordinary skill in the art may determine the optimal number of genes that
must be represented by nucleic acid fragments immobilized on the solid support to
10   effectively differentiate between samples that are at the various stages of stem cell
differentiation, including terminal differentiation, proliferating stem cells, stem cells
dedicated to a given differentiation pathway and/or stem cells with increased survival
rates. Preferably, at least about 5, 10, 20, 50, 100 , 150, 200, 300, 500, 1000 or more
preferably, substantially all of the detectable mRNA species in a cell sample or
15   population will be present in the gene expression profile or array affixed to a solid
support. More preferably, such profiles or arrays will contain a sufficient representative
number of mRNA species whose expression levels are modulated under the relevant
differentiation process, disease, screening, treatment or other experimental conditions. In
most instances, a sufficient representative number of such mRNA species will be about 1,
20   2, 5, 10, 15, 20, 25, 30, 40, 50, 50-75 or 100 in number and will be represented by the
nucleic acid molecules or fragments of nucleic acid molecules immobilized on the solid
support. For example, nucleic acids encoding all or a fragment of one or more of the
known genes or previously reported ESTs that are identified in Tables 2 and 3 may be so
immobilized. Additionally, the skilled artisan may select nucleic acids encoding the
25   protein cell surface markers discussed above at page 8 (*i.e.*, CD 34) in order to help
identify the particular stage of differentiation of a given stem cell population and to
identify agents that are involved in promoting such differentiation. The skilled artisan
will be able to optimize the number and particular nucleic acids for a given purpose, i.e.,
screening for modulating agents, identifying activated stem cells, etc.

In general, nucleic acid fragments comprising at least one of the sequences or part of one of the sequences of Table 2 can be used as probes to screen nucleic acid samples from cell populations in hybridization assays. Alternatively, nucleic acid fragments derived from the identified genes in Table 3 which correspond to the sequences of Table

5　2 may be employed as probes. To ensure specificity of a hybridization assay using probe derived from the sequences presented in Table 2 or the genes of Table 3, it is preferable to design probes which hybridize only with target nucleic acid under conditions of high stringency. Only highly complementary nucleic acid hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the

10　amount of complementarity which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the difference in stability between the probe:target hybrid and potential probe:non-target hybrids. .

Probes may be designed from the sequences of Table 2 or the genes of Table 3 through methods known in the art. For instance, the G+C content of the probe and the

15　probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available in Sambrook et al. (Molecular Cloning: A Laboratory Approach, Cold Spring Harbor Press, NY, 1989) or Ausubel et al. (Current Protocols in Molecular Biology, Greene Publishing Co., NY, 1995). Any available format may be used in designing hybridization assays, including immobilizing the probes to a solid

20　support or immobilizing the cellular test sample nucleic acids to a solid support.

It should be understood that the foregoing discussion and examples merely present a detailed description of certain preferred embodiments. It therefore should be apparent to those of ordinary skill in the art that various modifications and equivalents can be made without departing from the spirit and scope of the invention. All documents,

25　patents and references, including provisional patent application 60/056,861, referred to throughout this application are herein incorporated by reference.

-34-

What is Claimed Is:

1.   A method to identify an agent that modulates the expression of at least one stem cell gene associated with the differentiation process of a stem cell population, comprising the steps of:

5   preparing a first gene expression profile of an undifferentiated stem cell population;

preparing a second gene expression profile of a stem cell population at a defined stage of differentiation;

treating said undifferentiated stem cell population with the agent;

10   preparing a third gene expression profile of the treated undifferentiated stem cell population;

comparing the first, second and third gene expression profiles; and

identifying an agent that modulates the expression of a least one gene in undifferentiated stem cells that is associated with stem cell differentiation.

15   2.   A method to identify an agent that modulates the expression of at least one stem cell gene associated with the proliferation of a stem cell population, comprising the steps of:

preparing a first gene expression profile of a non-proliferating stem cell population;

20   preparing a second gene expression profile of a proliferating stem cell population;

treating the non-proliferating stem cell population with the agent;

preparing a third gene expression profile of the treated stem cell population;

25   comparing the first, second and third gene expression profiles; and

identifying an agent that modulates the expression of a least one gene that is associated with stem cell proliferation.

-35-

3.   A composition comprising a grouping of nucleic acid molecules that correspond to at least part of the sequences of Table 2 or genes of Table 3 affixed to a solid support.
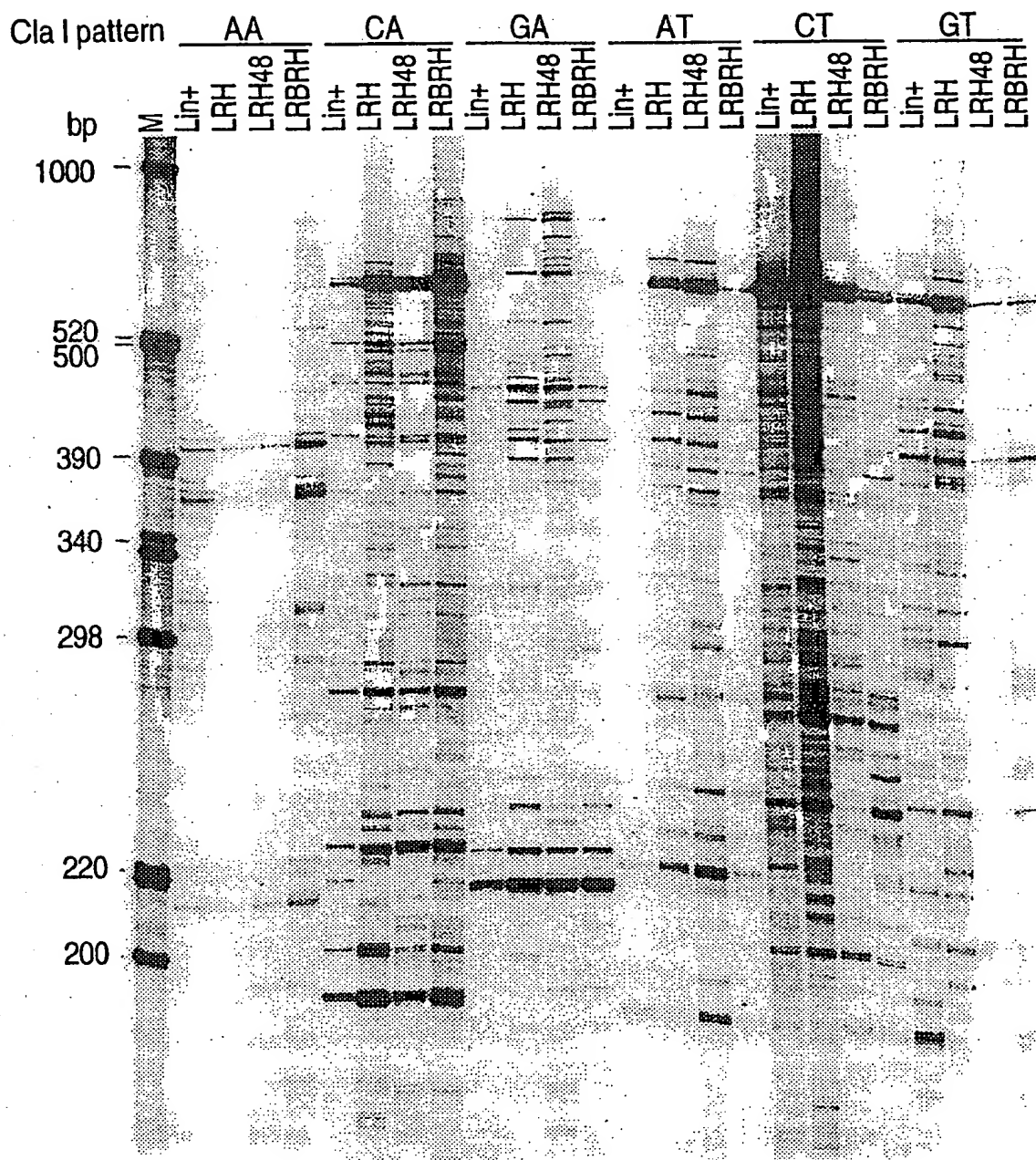
FIG. 1

FIG. 1 (Cont.)
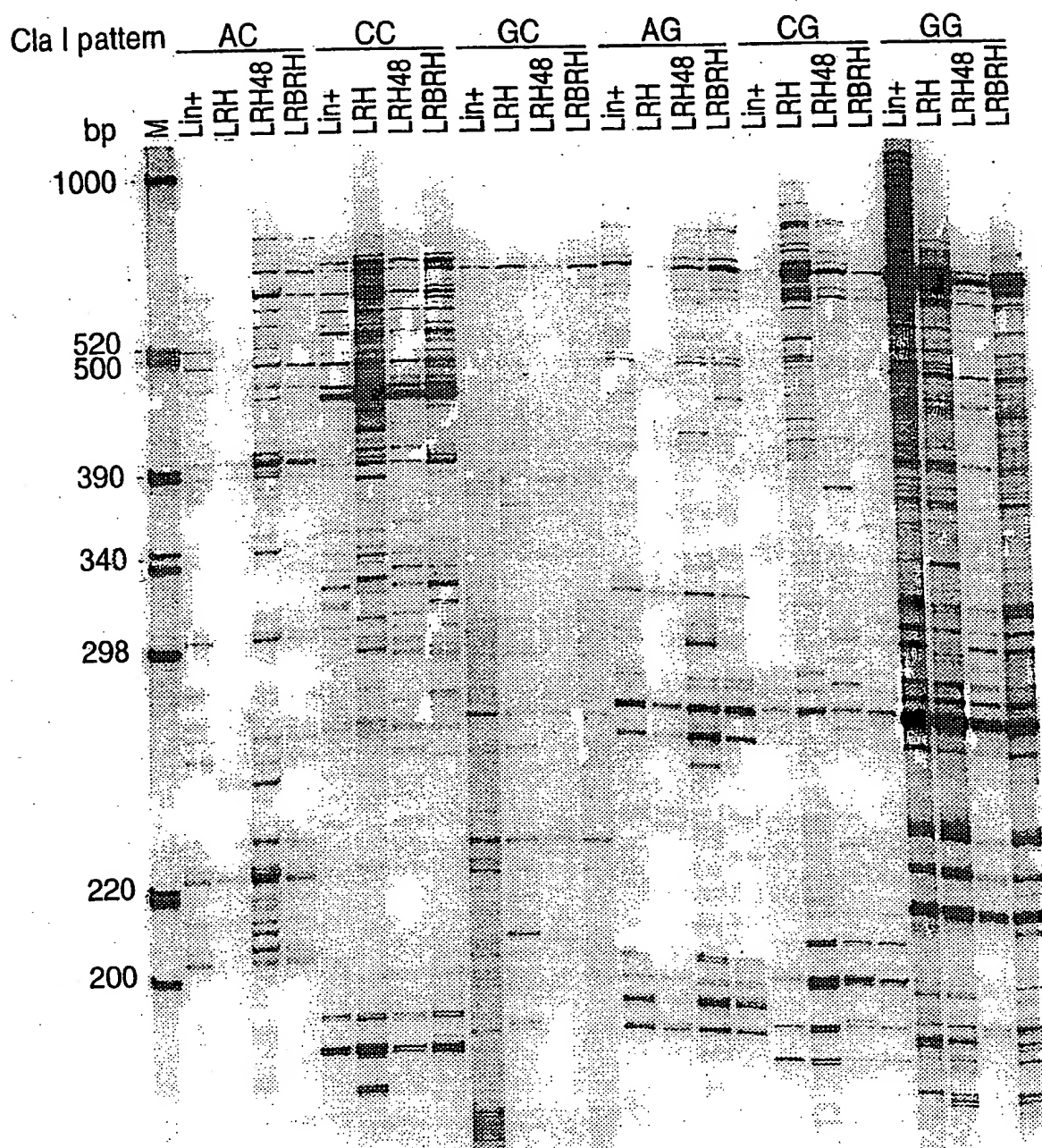
SEQUENCE LISTING

<110> Yale University

<120> A PROCESS TO STUDY CHANGES IN GENE EXPRESSION IN STEM
      CELLS

<130> 44574-5014-WO

<140> PCT/US98/17283
<141> 1998-08-21

<160> 93

<170> PatentIn Ver. 2.0

<210> 1
<211> 178
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 1
tttaattagc gctctatata cattgcggaa cttcccccga ctgcagcagt ttgactttgg 60
cacaacatca agttccattt cttttggaca ttggattctg ttttganagt atgtatgccc 120
caaagcattt tcagtgtcat caggattagt tgggcccatt cacagtaatt cananatc   178

<210> 2
<211> 148
<212> DNA
<213> murine

<400> 2
tagaatacct ggatggcttc tcttgtccac ccgatctccc gtgttaccaa tgtgtatggt 60
ctccttctcc cgaaagtgta cttaatcttt gctttctttg cacaatgtct ttggttgcaa 120
gtcataagcc tgaggcaaat aaaattcc   148

<210> 3
<211> 203
<212> DNA
<213> murine

<400> 3
gatctggcta gacagttatt ctgaactatg gcttcaagat gaacaagaca agcctaaaag 60
gatggagaga ggcaatggag ataatgtttt ggaggaagta tgtcactcaa gcatgaactc 120
tgtttattta gaaatgagat tccatatatg tggtacatgt ggaaagaatc taaaaagtcc 180
tttaaatttt ttcattccaa aag   203

<210> 4
<211> 336
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)

<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 4
ctnnannagc actcttcttg gccagacctc tgtccaaggc tcattagaaa gctgggggttn 60
tgtncacgtn acnnacttna tcnaaactnt tgctgtnttg gcataagttg tgtntctgga 120
ctgtnntgta ttcccctcta gacaaaggan caacnnaaaa gtnnttgcnn nctttnccag 180
aacatnctca aagcctntga tggaggagca caaggaccct gtctgctgag ggcccatggn 240
tcctctcagg ggtttctncc caccnaggca gtgccttcat tngctagtng tncagttact 300
tgtagnttat ctttnaataa atttnaataa aanctta   336

<210> 5
<211> 113
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 5
ctagattgtg tggtttgcct cattgtgcta tttgcgcact ttccttccct gaagaaatan 60
ctgtgaanct tctttctgtt cagtcctaan attcnaaata nagtgagact atg   113

<210> 6
<211> 164
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 6
ctcaagnacg ggccaggtaa gggcctttaa cacaactaaa tcaaggtgtg cttncctccg 60
ggttctatgc aagcaaggca tacacactgc actctcncnc tcnctaaact ggaaangtac 120
agtngcaggg ctggtttcag acnacgtgat gcntgtttac aaac   164

<210> 7
<211> 141
<212> DNA
<213> murine

<400> 7
tttttattca atatattaaa tatattaatc agaaaagtca catcctataa atccaggaaa 60
atacacaaat ataaatcaga atctgtcaat caccttcttg agtgacagtt atgtacacat 120
ggaaggagag cggaagagat c   141

<210> 8
<211> 224
<212> DNA
<213> murine

<400> 8
cgatatacac catcggtctg gggccaacgc taatactact tggtgctgcc aattgaattc 60
tggtttgctg tgaatctcta tcaacaagag tatcatttgt gaatgcttta atttattgag 120
aaagaacaag aagatgatgg atacattgat acatttgcgc agccttgcag cctgactcaa 180

ttctgctgtt catcagtttt aatgtccttt ctgtgtcata cgtg   *22r*

<210> 9
<211> 210
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 9
gatctttttt ccttcactta ttgctgaaac caagngcaca attcccatta agngaaggat 60
ctctgtgctg taaactaaac aaattgtgca ttttttctgg ggccattgtt tttggtttat 120
tttgttattt tgttttgttt ttgttttttt ggtttcattt tgttttgggt tggtccaatt 180
ttaaaaggaa atactacaat aaaaatgtta   210

<210> 10
<211> 163
<212> DNA
<213> murine

<400> 10
gatctgattt gctagttctt cctggtagag ttataaatgg aaagattaca ctatctgatt 60
aatagtttct tcatactctg catataattt gtggctgcag aatattgtaa tttgttgcac 120
actatgtaac aaaactgaag atatgtttaa taaatattgt act   163

<210> 11
<211> 176
<212> DNA
<213> murine

<400> 11
gcgatgttct tctactcaca actcacgttg gtggcctggg cctgaacttg actggagctg 60
acactgtggt gtttgtggag catgactgga accctatgcg agatctgcag gccatggacc 120
gggcccatcg tattgggcag aaacgtgtgg ttaatgtcta ccggttgata accaga   176

<210> 12
<211> 123
<212> DNA
<213> murine

<400> 12
gatctggaag ggaatgtcca aagagaagaa ggaggagtgg gaccgcaagg ctgaggatgc 60
taggagggag tatgagaaag ccatgaaaga gtatgaagga ggaagagggg actcatctaa 120
aag   123

<210> 13
<211> 196
<212> DNA
<213> murine

<400> 13
gatcttcgac acagagaagg agaaatacga gattacagag cagcgaaagg ctgaccagaa 60
agctgtggat ttgcagattt tgccaaagat taaagctgtt cctcagctcc agggctacct 120
gcgctctcag ttttccctga caaacgggat gtatcctcac aaaactggtct tctaaattgt 180
taacctaatt aaacag   196

<210> 14

SUBSTITUTE SHEET (RULE 26)

4

```
<211> 225
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 14
actcaatctc ttcaaactct ttatactggn ctatnatnag ngggatgtg  ncaanatnga  60
cnctggtggt gtatgaaaga aaagntcnat ggacntnggc atnccaagat tgaattcacc  120
tgcttcctac gatgtgtgaa actgctaata gcaaaatatc tctanggtta tgangagtac  180
tgtcgttctg caaatattca cttcanaact anncaccacg ttnaa                  225


<210> 15
<211> 244
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 15
ctagataatc ccttactgag tctttcttcn caggtgattc anttgagttg acaattannn  60
ctaagaattc aatggactan tgaggtgcct cagcagntaa tagcanttgc tgttcttcca  120
gaggaccaga gttcagtttc tcatcccaag ttgggctgct cgtnagtgtc ggtaantcca  180
gcttcagggg cttgaattta tactgaccat gggcacctgt accccaacac anacacatac  240
acat                                                               244


<210> 16
<211> 233
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 16
ctagaagtta atcctgtnaa gcatggtaag aatancattc tcaanatctt gagttaanaa  60
agatcttgga ggnggctggn gagatggctc antggttaag ancnctgact gctcttccag  120
aggtcctgan ttcaattccc ancaaccaca tggtggntca caaccanctg taatgatacc  180
tgatgccatc ntccgtggtg tatctgaana canctacagt gacagctaca ncg         233


<210> 17
<211> 260
<212> DNA
<213> murine

<400> 17
ggattttatt ctaggcttgg ccagatacag gttggcatcc taggggagga agataacaat  60
gtcataggtg aatttgttag gagaggcaag acatgggaaa tcattgattt cttcagattt  120
ctttaaagca aattagaaga taaatgtcta aaagagatac acttaaaaaa tggtgaaact  180
ataacccctt aaggagagcc agatgtggca ggagccaggt ctgaaaatgg tagctgaagt  240
```

aagcagacca gcgtaagatc   260

<210> 18
<211> 154
<212> DNA
<213> murine

<400> 18
cgatgagtca gagaggaagt ggacagtgcg ttattcatta cagcaaagga tttcgttggc 60
atcaaaatct aagtttgttt tacaaagatt gttttagta ctaagctgcc ttggcagttt 120
gcatttttga gccaaacaaa aatatattat tttc   154

<210> 19
<211> 340
<212> DNA
<213> murine

<400> 19
cgattcaatt gtataaatga ttataatttc tttcatggaa gcatgatcct tctgattaag 60
aactgtaccc catatttat gctggttgtc tgcaagcttg tgcgatgatg ttatgttcat 120
gttaatccta tttgtaaaat gaagtgttcc tgaccttatg ttaaaaagag agaagtaaat 180
aacagacatt attcagttat tttgtccttt atcgaaaaac cagatttcat ttttcctttt 240
tgtttgtgat ctcatttgga ataattggc aagttgaggt actttcttcc catgctttgt 300
acaatataaa ctgttatgcc tttcagtgcg ttactgtggg   340

<210> 20
<211> 277
<212> DNA
<213> murine

<400> 20
ctagaggtgg gaactggctc cactccacac agcagccagt tagttagtga cggtcagctg 60
catgcagggg aatgaaggac tcggagagaa cgttctgtgc tatgtgtgtt ccatagagat 120
taaaaaggag gcctggagcc gagcatggtg gtgcacgcct ttaatcccag cacttgggag 180
gcagagtcag gtggatttct gagttcattg ccagcctggt ctacagagtg aattccagga 240
caggcagggc tacacagaga aaccctgtct caaaaaa   277

<210> 21
<211> 66
<212> DNA
<213> murine

<400> 21
ctagaatttg cagtagcatt aattcaagcc tacgtattca ccctcctagt aagcctatat 60
ctacat   66

<210> 22
<211> 121
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 22
ctagacataa gatattgtac ataaaganaa ttttttttgc ctttaaatag ataaaagtat 60
ctatcagata aaaatcangt tgtaagttat attgaagaca atttgataca taataaaaga 120
t   121

```
<210> 23
<211> 127
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 23
ggggagnnnn cnagnaanna gantcgtacg taaanagaan nntggtgcnt ttanatagaa 60
aangtactat canataanaa tcaggttgta agttatattg aagacgnttt gatacataat 120
aaaagat                                                          127

<210> 24
<211> 105
<212> DNA
<213> murine

<400> 24
ctagactgac aaagactttt tgtcaacttg tacaatctga agcaatgtct ggcccacaga 60
cagctgagct gtaaacaaat gtcacatgga aataaatact ttatc                105

<210> 25
<211> 85
<212> DNA
<213> murine

<400> 25
ctctcttgcc acccagatgg ttaggatgat tctgaagatg atgacatccg taagcctgga 60
gaatctgaag aataaactgt accat                                       85

<210> 26
<211> 85
<212> DNA
<213> murine

<400> 26
ctctcttgcc acccagatgg ttaggatgat tctgaagatg atgacatccg taagcctgga 60
gaatctgaag aataaactgt accat                                       85

<210> 27
<211> 316
<212> DNA
<213> murine

<400> 27
gatctcggaa tggacccaac tgctcctgct ccaccggcgg ctcctgcact tgcaccagct 60
cctgcgcctg caagaactgc aagtgcacct cctgcaagaa gagctgctgc tcctgctgtc 120
ccgtgggctg ctccaaatgt gcccagggct gtgtctgcaa aggcgccgcg gacaagtgca 180
cgtgctgtgc ctgatgtgac gaacagcgct gccaccacgt gtaaatagta tcggaccaac 240
ccagcgtctt cctatacagt tccaccctgt ttactaaacc cccgttttct accgagtacg 300
tgaataataa aagcct                                                316

<210> 28
<211> 136
<212> DNA
<213> murine
```

<400> 28
attcagacga atgagactcc tccacattgg agacaagaga tgcagagagc tcagagaatg 60
agggtgtcaa gtggtgaaag atggatcaaa ggggataaga gtgagttaaa tgaaataaaa 120
gaaaatcaaa ggagcc   136

<210> 29
<211> 243
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 29
ngcnnnnnnn ccagnaggag gagaagatga ctggccagta tcanaatggg ataagatgag 60
gcgcgccctg gagtacacca tctacaacca ggagctcaac gagacgcgcg ctaagctcga 120
cgagctttct gctaancgag aaacnagtgg agagaaatcc ngacaactaa gggatgccca 180
gcaggatgca ngagacaaaa tggaggatat tgagcgccag gttagagaac tgaaaacaat 240
nat   243

<210> 30
<211> 359
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 30
ctcaaggaaa agacagcacc ncgtgcctgg catctgntgn nttagntnat ntnnaantnt 60
cnnntngncc tggcaacggt tcctgaacna attaccactc cttcttgcca gtcnaanagg 120
gtgggaaagt ccgagcctta ngacccagtt tcagttctgg tttcttccct cctgancacc 180
atcggttgtt agttgccttg agttgggaac gtttgcatcg acacctgtaa atgtattcat 240
tctttaattt atgtaaggtt ttntgtnctc aattctttaa gaaatgacaa attttggttt 300
tctactgttc aatgagaaca ttaggcccca gcaacacgtc attgtgtaaa naaataaaa 359

<210> 31
<211> 139
<212> DNA
<213> murine

<400> 31
cgatggctcc atcctggcct cactgtccac cttccagcag atcggctcag caagcaggag 60
taggatgagt ctggccccctc catcgtgcac cgcaaatgct tctaggcgga ctgtttttaca 120
cccttttcttt gacaaaacc   139

<210> 32
<211> 354
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)

<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 32
cnnatgctac atgctgnagg atgcctaagg ctgcccccca ccatcccctg gctctgctgn 60
ccggancaaa ttgcttccag atgtgacttt ggaaccttcn cacccctnac ccnaccnntc 120
tcnagaannt cttttatta aaggaggaaa nannacatcc aagaaaangg ggggaggggg 180
gatggaaann cgcatcccct ttctagccag ctgttcccaa aaggtaccct tcctctctgc 240
tgctccccaa acncaaancc cacttcngan cctccaccta aancatcang caagtcacnt 300
acaccctgtt tanccccna ctctctgctt ataccnggga acaattnntg ctcg   354

<210> 33
<211> 412
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 33
cgatggtggg gatcttactg gggaagagga aggaccatta gcacaccatc atgatgtcag 60
atgacaaaat ggaagccaag acaccttgaa ggtgactttc taggaaggtc ttaagcatgt 120
aatgtccctt tatcagaggg aaggggacaa actcagggca gccctgtcca ggtagaaata 180
tttttgcccc cctgtctgat gttgatgagg ggtcatacca nccagggaga ccctctggga 240
ggaagctgcc acacacaang actctggaag tatccagatg tgagcccagc cagggtccta 300
tggttccaaa tctgaanaaa aggttttca cacactcctt gctttctgct aagataanaa 360
aggcgtcact ctgccagagt gtgactttt acagattaaa taaagctgtt at   412

<210> 34
<211> 239
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 34
gatctactcc attcccctgg aaatcatgca gggcaccggg ggtgagctgt ttgatcacat 60
tgtctcctgc atctccgact tcctggacta catggggatc aaaggccccg gatgcctctg 120
ggcttcacct tctcgtttcc ctgcaagcag acgagcctat attgcggaat cttgatcacg 180
tggacaaagg gattcaaagc caccgactgt gtgggtcacn atgtanccac tttactgag 239

<210> 35
<211> 93
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 35
gatctgagtt cgaggccagc ctggtctaca gagtgagttc caggncagcc aggnctacac 60

agagaaaccc tgtctcgaaa aaacagaaag aga   93

<210> 36
<211> 130
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
       throughout the sequence may be A, T, C or G

<400> 36
ctttcattaa aaagaaacca ggggctggan agatggctca gtggttaaga gcaccaactg 60
ctcttcccga aggtcctaag ttcaaatccc agcaaccaca tggtggctaa caaccactcg 120
taatgagatc   130

<210> 37
<211> 234
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
       throughout the sequence may be A, T, C or G

<400> 37
atcgcntggc tctcctgngg cctggcntac gacnngaaaa ggagtgtcca cggctgctgt 60
cgnggccacg attaattaaa actgaagtac cgaggntncc ccagngncng antgtggggt 120
cnngccnttc ntgntccaca anccaacttg gcagacgctt actgtnctgt caactntcnn 180
nngaataccn ccacccncat gctaaaatga tgactgacgt taanccatgc tggt   234

<210> 38
<211> 251
<212> DNA
<213> murine

<400> 38
cgatgacaaa ggagtcctga ggcagattac tctgaatgac cttcctgtcg gaagatcagt 60
ggacgagaca ctgcgtttgg ttcaagcctt ccagtacact gacaagcatg gagaagtctg 120
ccctgctggc tggaaacctg gtagtgaaac aataatccca gatccagctg gaaaactgaa 180
gtatttcgac aagctaaact gaaaagtact tcagttatga tgtttggacc ttctcaataa 240
aggtcattgt g   251

<210> 39
<211> 179
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
       throughout the sequence may be A, T, C or G

<400> 39
cgatgctgaa taagctcctc aaaaagtggt aaatttaacc ttttnaaaaa acaagctttc 60
tctgtacagc tctggctgtt ttgttctgga atacattctg tagaattgtc tggcctctaa 120

```
cttggagatc caactccctc tgcctcttga gtgctgggat taatggcatg tgacactgt    179
```

```
<210> 40
<211> 219
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 40
cgatgacctc atgccggccc agaagtgaag cctggccctc gccaccatca ggctgccgct    60
tcctaactta ttaaccgggc agtgcccgcc atgcatcctt gangtttgcc gcctggcggc    120
tgagccctta gcctcgctgt agagacttct gtcgccctgg gtagagttta ttttttgat    180
ggntaanctg ttgctgacac tgaaaataan ctagggttt                          219

<210> 41
<211> 303
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 41
cgatcaatga aaagatgacg agtttctttc aaatgggcag ttactccctg ataacttcat    60
agctgcctgc acagagaaga aaatccctgt tgtgtttaga ctacaagagg gttatgatca    120
tagctactac ttcattgcaa ctttcatcgc tgaccacatc agacaccatg ctaagtacct    180
gaatgcatga naagcctcag ccaagagaat ctcatcagga ggccggaagg gaatcaacag    240
gagtgctgac ttcctcgcag aagatcatgc tcctgcagct gaatcgcttt tctgaataaa    300
tat                                                                 303

<210> 42
<211> 460
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 42
cgatgtntac ttcattgcca ccctgtcant cctctggaag gtgtccgtca tcaccttggt    60
cagctgtctc ccctctatg tcctcaagta cctgcggaga cggttctccc cacccagcta    120
ctcgaagctc acttcctaag ctgcagggct gcctcgggca gggcctccgg cctccggcgc    180
tctcccagga ggaggtcaag ttccacacgc acgagccgcc tctgctggac ggtgcagtca    240
tggctggcac atgaggcttc gctgaggcga cactgggcac ctaatgggga tggaacattg    300
gtggaaccgg agggagggac ctgagagctg tacctatcag aaccttgggt gctaagctgt    360
gcttgagggg aagacgtggg accggatggc ccgtctgagg tttgtggggt cactgtgcaa    420
gcttccttat ggtttgaacc tcttgtcatg tgataaaagt                         460

<210> 43
<211> 120
```

```
<212> DNA
<213> murine

<400> 43
cgatttacgt atttgactga aatgaaagtt ccactaaacg gtatttgctc ttgtgatatg 60
tggcacattg tgatattttc ttagtctgtt ctgtttcatt taaaaaataa aactgctgat 120


<210> 44
<211> 132
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 44
ccgatgtncg ataatagtaa ataccttaat tanttaaata attcattgna ttgtttcaga 60
gacgtttgga aattactgta tacatttaca acctaatgac ttttgtattt tattttttcaa 120
aanaaaagct ta                                                     132


<210> 45
<211> 240
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 45
cnttngnnnn tccntncatc ncngcngtnt gagtcccncc caannagtcc atccaananc 60
canngcatnn cagctttatc atgacaacaa antggagnaa gaagaagatg agtttcggcc 120
actgttgagg caaatcnntg nnnantcnta atanacacct ggtccgctca tccttcaacg 180
ttgtttntnta naanttacct cccagtagaa angctagcaa ntttnacctg ccacnggttn 240


<210> 46
<211> 126
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 46
cgatcagatg tcacgcggga cacancnccg ccncagtnaa tggnaatata tttgcatgtt 60
accccaaatt ancttctntg catngaacat angtangtgt ctttggggac acgtgtgttc 120
tactac                                                            126


<210> 47
<211> 383
<212> DNA
<213> murine
```

12

```
<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 47
cgatttacaa atgaacaanc aagattacat atantgaaaa tccacgcagg acctattaca 60
nagcatggtg aaatagatta tgaagcaatt gtaaagcttt cagatggctt taatggagca 120
tgacctgaca aatgtttgta ctgaagcagg tatgtttgca attcgtgccg atcatgattt 180
tgtanttcag gaagacttca tgaaagcagt cangaangtg gctgactcca agaagctgga 240
gtccaagctg gactacaaac ctgtgtgatt cactannagg gtttggtggc tgcatgacag 300
acattggttt aatgtanact taacngttan ngaaactaat gtanntattg gcaatganct 360
tattanaagt gaatanacat gtg                                          383

<210> 48
<211> 255
<212> DNA
<213> murine

<400> 48
cgatgttttt aattaagaag aaattcactt tctcattacc tatgaatctg tgccagggca 60
ggtgattttt gagtatgaga actttgtcct ctccacagtt gtcacaaaaa tggttccttc 120
tcattgaact attgtggcat gctaattaag aagtgagtga ccacttggga ggcagaggca 180
ggtggatttc tgagtttgag gccagcctgg tctacaaagt gagttctaag acagccaggg 240
ctatacagag aaacc                                                   255

<210> 49
<211> 243
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 49
ccaagnaata tggtctaatc aaaggtcgtc tgtctgcttt tgattgtcta catcacagca 60
atccctggga atttctatcc attttaaatg cngccgcttt catctgttta gccagcacac 120
ccaatggttt cactaactag cccagttgac cttttggaag tttgagcctt gagcaccttc 180
aacaaaattg agcactctga ttaggatatc cactttgcaa ataaaaccaa atgttttgtc 240
aac                                                                243

<210> 50
<211> 358
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 50
cgatgagggg aagatgacct gggccgggga ggccatccct tatccaagat cacagggaat 60
tctgggaaga ggttggcctg tggcatcatt gcacgctctg ccggcctttt ccagaacccc 120
aagcagatct gctcctgtga tggcctcact atctgggagg agcgaggccg gcccattgcc 180
ggtcaaggcc gaaaggactc agcccaaccc ccagctcacc tctaaacaga gcctcatgtc 240
```

```
aggttatttg gtcctcgtag ctgaacatct tcttgcagag ggagctgcng gcccttgctt 300
gtacaggcct aagtacaggg cagataagtg ctgtagcctg aacaaattaa attgttac   358
```

<210> 51
<211> 355
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
        throughout the sequence may be A, T, C or G

<400> 51
```
cgattagctg nggtctctag ganatactcg tcactatatg agctcaggan gccagctctt 60
agtagctctg aancaggtga agaatcctcc tctgaggaaa cagactggga ggaagaagca 120
gcccattacc agccagctaa ttggtcaaga aaaaagccaa aagcngctgg cgaaagtcag 180
cgtactgttc aacctcccgg cagtcggttt caaggtccgc cctatgcgga gcccccgccc 240
tgcgtagtgc gtcagcaatg cgcagagggg caatgcgcag agaggtgcgc agagggcag 300
tgcgcagaga ggtgcgcaga gaggcagtgc gcagagaggc agtgcgcaga ctcat     355
```

<210> 52
<211> 213
<212> DNA
<213> murine

<400> 52
```
cgatttctaa atcagtctcg cctgtgctag gatgaccggt aatgagcctg tttaaaataa 60
gacttaaaag tgtcgtgcgt tggccgggcg gtaggggcgc atgcctttaa tttcataact 120
tggaggtaga gacaggcgga tctttgtgag ttcaaggtca gcctggtgta cagagtgact 180
tccagaacag ccagggctgt taaacagaga aac                             213
```

<210> 53
<211> 113
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
        throughout the sequence may be A, T, C or G

<400> 53
```
ttgttttgtt nttcagatag ggtcttacat atcccatgct ggtctcaaac tcacattatg 60
catgcgggga aagccattta ctgactgata tacccctggc cctaagatag atc        113
```

<210> 54
<211> 108
<212> DNA
<213> murine

<400> 54
```
cgatcgtcgt tctggtaaga agctggaaga tggccccaag ttcctgaagt ctggccattt 60
aagtttaata gtaaaagact ggttaatgat aacaatgcat cgtaaaac            108
```

<210> 55
<211> 257
<212> DNA
<213> murine

```
<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G


<400> 55
cgatcgtcgt tctgagtaan aagctggaan anggccccaa gttcctgnng tctggcgatg 60
ctgccattta agttnannag ananaagact ggctnatgat aacaatgcan cntaaaacct 120
tcaggnaggn aacgaatgtt gtggaccatt ttttntgngt gtggcagttt naagttatna 180
agntttcaaa ancantactt nttaanggga acaacttgac ccatcanctg tcacagaatn 240
ttgangacca ttaacac                                                257


<210> 56
<211> 151
<212> DNA
<213> murine


<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G


<400> 56
nctacgatca tctagatcta ctagacctac nacnagacca tgggccaaan atggtcgacc 60
tgcaaacttg caaggtttat tttanataca cattatggcg ttttatnttt tgtaattcta 120
agttgtaatt cagcttttaa caaatctttt t                                151


<210> 57
<211> 152
<212> DNA
<213> murine


<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G


<400> 57
ccaagnanat cnagactact agacctacta cnagaccatn ggncaaacat ggtcgaccnn 60
caaacgnata ngtatatttn anatacacan anatagcgtt ntatgtctng taattctaag 120
tngtanatca nctattanca aaatctttnt tt                               152


<210> 58
<211> 188
<212> DNA
<213> murine


<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G


<400> 58
cgatggaagt tctgctgagc ccttctgacg taaccctggc natggctaac actgtccttc 60
ctgcaatgtt cntggtggac acancttctc tgganatacc ctgaangtgg cacgccctgt 120
tccagcccac ctggtgtgca cttttgccc tctttacctc attantaaat gttttcntgc 180
```

SUBSTITUTE SHEET (RULE 26)

tcctaatg   188

<210> 59
<211> 136
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 59
ctnagnaagg anctgtactt cgtattgcaa ggcagtctct tgtgtcttct tagagtgtct 60
tccccatgca cagcctcagt ttggagcact agtttataat gtttattaca attttttaata 120
aattgantag gtagta   136

<210> 60
<211> 365
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 60
tcntcnttct ggtaagaact ggaatatggc cccaagttcc tgaagtctgg cgatgctgcc 60
attgttgata tggtccctgg caancccatg tgtgttgaga gcttctctga ctaccctcca 120
cttggtcgct ttgctgttcg tgacatgagg cagacagttg ctgtgggtgt catcaaagct 180
gtggacaaaa angctgctgg agctggcnaa gtcaccaagt ctgcccanaa agctcagaag 240
gctaaatgaa tattacccct aacanctgcc accncantct taatcagtgg tggaagaacg 300
gtctcagaac tgttngtctc aantggccat ttaagtttaa tantaaaaga ctggttaatg 360
ataac   365

<210> 61
<211> 357
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 61
cgatcntcgt tctggtaaga nncnggaaca tggccccaag ttccngannt ctggcgangc 60
ngccantgtt gatatggtcc ctggcaagcc catgtgtntt gagagcttca cnnacnaccc 120
tccanttggt cgctttgctg ttcgtgacat gaggcagaca gttgctgtgg gtgtcancaa 180
anctgtggac aananggctg ctggagctgg caagntcacc aantctgccc agaaagctca 240
gaatgctaaa tnaatattac ccctaanacc tgccacccca gtcntaatca gtggtggaat 300
aacngtctca gaactgtttg tcncaattgg ccanttangt ttaatnatac aagactg 357

<210> 62
<211> 305
<212> DNA
<213> murine

```
<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 62
gnnnnnnnnn nncnangaaa aagaggtgaa aaatgcttgg ctctagctga tgacagaaag 60
ctgaaatcca tcgccttccc atccattggc agcggcagga acgggttccc ggaagcagac 120
agcggcccag ctcattctga agtgccatct ccagctacnt tgtctccacg atgtcctcct 180
ccatcaaaac tgtgtacttc atgctttttg acagtgagag cataggtatc tatgtgcagg 240
aaatggccaa gctggacgcc aactaggcca gtgatcccta gagccagcac atgcggtgtc 300
cccca                                                          305

<210> 63
<211> 327
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 63
ctnangaaag ctgctggggc nccctgacat cactcatcac tcactatgct accaattcta 60
tttatttcgg aattacaaga tatcgggaat ctctctgcag gctggactgg caggctgtgg 120
ggtgggcggg acacggctct taacatttnc agagggaaac gcgcanatgt ccaaaagtct 180
aaataaatgc attcagaggt ttntggggtc catggccaag tggagttccc ccncagggg 240
aggtggggta agtgcctcca ggaaggcagg cagcctgcct tanacttgca ncccggntgt 300
gggaatgaat cattggagta ataaact                                  327

<210> 64
<211> 271
<212> DNA
<213> murine

<400> 64
cgatgccaat ggcatcctca atgtttctgc tgtagataag agcacaggaa aggagaaagt 60
ctgcaaccct atcattacca agctgtacca gagtgcaggt ggcatgcctg ggggaatgcc 120
tggtggcttc ccaggtggag gagctccccc atctggtggt gcttcttcag gccccaccat 180
tgaagaggtg gattaagtca gtccaagaag aaggtgtagc tttgttccac agggacccaa 240
aacaagtaac atggaataat aaaactattt a                             271

<210> 65
<211> 310
<212> DNA
<213> murine

<400> 65
cgatgaagat gaggtcactg cagaggagcc cagtgctgct gttcctgatg agatcccccc 60
tctggaaggc gatgaggatg cctcgcgcat ggaagaggtg gattaaagcc tcctggaaga 120
agccctgccc tctgtatagt atccccgtgg ctcccccagc agccctgacc cacctggatc 180
tctgctcatg tctacaagaa tcttctatcc tgtcctgtgc cttaaggcag gaagatcccc 240
tcccacagaa tagcagggtt gggtgttatg tattgtggtt tttttgtttg ttttattttg 300
ttctaaaatt                                                     310

<210> 66
<211> 579
```

```
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
       throughout the sequence may be A, T, C or G

<400> 66
cgatgccaat ggcatcctca atgtttctgc tgtagataag agcacaggaa aggagaacaa 60
gatcaccatc accaatgaca agggccgctt gagtaaggaa gatattgagc gcatggtcca 120
agaagctgag aagtacaagg ctgaggatga gaagcagaga gataaggttt cctccaagaa 180
ctcactggag tcctatgcct tcaacatgaa agcaactgtg gaagatgaga aacttcaagg 240
caagatcaat gatgaggaca aacagaagat tcttgacaag tgcaatgaaa tcatcagctg 300
gctggataag aaccagactg cagagaagga agaatttgag catcagcaga aagaactgga 360
gaaagtctgc aaccctatca ttaccaagct gtaccagagt gcaggtggca tgcctggggg 420
aatgcctggt ggcttcccag gtggaggagc tcccccatct ggtggtgctt cttcaggccc 480
caccattgaa naggtggntt aagtnatcca nnaagaaagg ntnccttttt ttccaaaggg 540
anccaaaaaa gtaanatgga taataaaacc tatttaatt                        579

<210> 67
<211> 186
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
       throughout the sequence may be A, T, C or G

<400> 67
cgatgccaat agnancccaa ntntctgcng tngataagac acangaaaag agaacaagat 60
caccatcacc aatgacaagg gccgcttgag taaggaagat attgagcgca tggtccaaga 120
tcaatgatga ggacaaacag aagattcttg acaagtgcaa tgaaatcatc agctggctgg 180
ataaga                                                            186

<210> 68
<211> 321
<212> DNA
<213> murine

<400> 68
cgattagcgg aggtctctag gagatactcg tcactagatg agctcaggaa gccagctctt 60
agtagctctg aagcaagtga agaatcctcc tctgaggaaa cagactggga ggaagaagca 120
gcccattacc agccagctaa ttggtcaaga aaaaagccaa aagcggctgg cgaaagtcag 180
cgtactgttc aacctcccgg cagtcggttt caaggtccgc cctatgcgga gcccccgccc 240
tgcgtagtgc gtcagcaatg cgcagagggg caatgcgcag agaggcagtg cgcagagagg 300
cagtgcgcag actcattcat t                                          321

<210> 69
<211> 321
<212> DNA
<213> murine

<400> 69
cgattagcgg aggtctctag gagatactcg tcactagatg agctcaggaa gccagctctt 60
agtagctctg aagcaagtga agaatcctcc tctgaggaaa cagactggga ggaagaagca 120
gcccattacc agccagctaa ttggtcaaga aaaaagccaa aagcggctgg cgaaagtcag 180
cgtactgttc aacctcccgg cagtcggttt caaggtccgc cctatgcgga gcccccgccc 240
```

```
tgcgtagtgc gtcagcaatg cgcagagggg caatgcgcag agaggcagtg cgcagagagg 300
cagtgcgcag actcattcat t                                            321
```

<210> 70
<211> 495
<212> DNA
<213> murine

<400> 70
```
gatctttgta ggcacaaaat gaatcccgca cctggtgacc catgatgctc gtactattcg 60
gtaccctgat cccctcatca aggtgaacga caccattcag attgatttgg agacaggcaa 120
aataactgac ttcatcaagt ttgacactgg gaacctgtgt atggtgactg gaggtgctaa 180
cttgggaaga attggtgtaa tcaccaacag agagagacat cccggctctt ttgatgtggt 240
tcatgtgaaa gatgccaatg gcaacagctt tgccactcgg ctgtccaaca ttttttgttat 300
tggcaagggt aacaaaccat ggatctctct tcccagagga aaaggaatcc gcctcaccat 360
tgctgaagag agagacaaga ggcttgcggc caaacagagc agtgggttga aatggtctcc 420
taggagacat gcctggaaag ttgttttgta caacctttcc taggcaacat acattgctag 480
aattaaacag ccatg                                                   495
```

<210> 71
<211> 136
<212> DNA
<213> murine

<400> 71
```
cgatcgagag ggcaaaccac ggaaggtggt tggttgcagt tgcgtagtgg ttaaggacta 60
tggcaaagaa tctcaggcca aggatgtcat cgaggaaata cttcaagtgc aagaaataaa 120
taaattttgg ctgatt                                                  136
```

<210> 72
<211> 140
<212> DNA
<213> murine

<400> 72
```
attccagatg aggaccacaa gcgactcatt gatttacata gtccttctga gattgttaag 60
cagattactt ccatcagtat tgagccggga gttgaggttg aagtcaccat tgcagatgcc 120
taagacaact gaataaatcg                                              140
```

<210> 73
<211> 216
<212> DNA
<213> murine

<400> 73
```
gatctataca gtcgggaaac gcttcaagga agcaaataac ttcctgtggc ccttcaagtt 60
atcttcccca cgaggtggga tgaagaaaaa gacaactcac tttgtagaag gtggagatgc 120
tggcaacagg gaagaccaga taaacaggct tattagacgg atgaactaag gtgtcaccca 180
ttgtattttt gtaatctggt cagttaataa acagtc                            216
```

<210> 74
<211> 151
<212> DNA
<213> murine

<400> 74
```
cgatgtggcc aaagtcaata ccctgataag gcccgacgga gagaagaagg cgtatgttcg 60
cttggctcct gattatgatg ccctagatgt tgccaacaag attgggatca tctaaactga 120
gtccagatgg ctaattctaa atatatactt t                                 151
```

<210> 75

```
<211> 90
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 75
gatctggaac catagatgcg agcatcagca acagaataca agaaatggaa gngngaatct 60
caggtgcaga agnttccata gagaacatcg                                   90

<210> 76
<211> 257
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 76
gcgatgcaaa atccttaata naattcttgc taaccgaatc caagaacaca ttaaagcaat 60
catccatcct gaccaagtag gttttattcc agggatgcng ngatggttta atatatgaaa 120
atccatcaat gtaatccatt ntataaacaa nctcaangac anaaaccaca tgatcatctc 180
gttagntgca gaaaaagcat ttgacaagat ccaacacaca ttcgtgataa nagttttggn 240
aagatcagga attcaag                                                 257

<210> 77
<211> 200
<212> DNA
<213> murine

<220>
<221> variation
<222> (various)
<223> bases designated as "n" at various positions
      throughout the sequence may be A, T, C or G

<400> 77
cgatnnaccc gctctacctc accatctctt gctaattcag cctatatacc gccatcttca 60
gcaaaccctta aatnaggtat taaagtaagc atcnagaatc anccatactc aacgtnacgt 120
caaggtgtac ccaatgnaat gggaagaaat gggctacatt ttcttatana agaacattnc 180
tatacccttt ntgaaactaa                                              200

<210> 78
<211> 56
<212> DNA
<213> oligo used in gene expression

<400> 78
acgtaatacg actcactata gggcgaattg ggtcgacttt tttttttttt tttttv      56

<210> 79
<211> 21
<212> DNA
<213> oligo used in gene expression
```

```
<400> 79
cttacagcgg ccgcttggac g                                    21

<210> 80
<211> 15
<212> DNA
<213> oligo used in gene expression

<400> 80
agcggccgct gtaag                                           15

<210> 81
<211> 30
<212> DNA
<213> oligo used in gene expression

<400> 81
gcggaattcc gtccaagcgg ccgctgtaag                           30

<210> 82
<211> 21
<212> DNA
<213> adapter oligo

<400> 82
cttacagcgg ccgcttggac g                                    21

<210> 83
<211> 15
<212> DNA
<213> adapter oligo

<400> 83
gaatgtcgcc ggcga                                           15

<210> 84
<211> 25
<212> DNA
<213> adapter oligo

<400> 84
tagcgtccgg cgcagcgacg gccag                                25

<210> 85
<211> 29
<212> DNA
<213> adapter oligo

<400> 85
gatcctggcc gtcggctgtc tgtcggcgc                            29

<210> 86
<211> 30
<212> DNA
<213> primer

<400> 86
gcggaattcc gtccaagcgg ccgctgtaag                           30

<210> 87
```

```
<211> 40
<212> DNA
<213> primer

<400> 87
ctctcaagga tcttaccgct tttttttttt tttttttat                    40

<210> 88
<211> 40
<212> DNA
<213> primer

<400> 88
taataccgcg ccacatagca tttttttttt tttttttcg                    40

<210> 89
<211> 40
<212> DNA
<213> primer

<400> 89
cagggtagac gacgctacgc tttttttttt tttttttga                    40

<210> 90
<211> 19
<212> DNA
<213> primer

<400> 90
tagcgtccgg cgcagcgac                                          19

<210> 91
<211> 19
<212> DNA
<213> primer

<400> 91
ctctcaagga tctaccgct                                          19

<210> 92
<211> 20
<212> DNA
<213> primer

<400> 92
cagggtagac gacgctacgc                                         20

<210> 93
<211> 20
<212> DNA
<213> primer

<400> 93
taataccgcg ccacatagca                                         20
```

SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12N 15/12
US CL : 435/6; 536/23.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Medline, WPIDS
search terms: hematopoietic stem cell, differential display

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TAGOH et al. Molecular Cloning and Characterization of a Novel Stromal Cell-Derived cDNA Encoding a Protein That Facilitates Gene Activation of Recombination Activating Gene (RAG)-1 in Human Lymphoid Progenitors. Biochem. Biophys Res. Commun. 1996, Vol. 221, pages 744-749, especially page 744.	1, 2
X	MOREB et al. Human A1, a Bcl-2-related gene, is induced in leukemic cells by cytokines as well as differentiating factors. Leukemia. July 1997, Vol. 11, Number 7, pages 998-1004, especially page 998.	1, 2

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 NOVEMBER 1998	**24 DEC 1998**

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JOHN S. BRUSCA
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

Box I   Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 3
   because they relate to parts of the international application that do not comply with the prescribed requirements to such
   an extent that no meaningful international search can be carried out, specifically:

   No sequence listing or computer readable form of sequence listing has been supplied, and claim 3 is drawn to specific
   sequences that therefore cannot be searched.

3. ☐ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II   Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable
   claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment
   of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers
   only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is
   restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest   ☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)★